

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:34:53 ; Search time 20.66 Seconds
(without alignments)
654,785 Million cell updates/sec

Title: US-09-462-845-2

Perfect score: 3489

Sequence: 1 MKKLITADDITAVSVTDPO.....HPQRIRKLNLYISSWFDQHL 657

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*

5: /cgn2_6/ptodata/2/iaa/PCT05_COMB.pap.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	1104.5	31.7	632	4	US-09-016-080-1
2	635	18.2	622	2	US-08-664-646A-2
3	635	18.2	622	2	US-09-066-285-2
4	635	18.2	622	3	US-09-261-006-2
5	635	18.2	622	4	US-08-951-088-2
6	272	7.8	755	5	PCT-US93-07923-3
7	272	7.8	759	5	PCT-US93-07923-2
8	272	7.8	766	1	US-08-230-491A-3
9	272	7.8	766	1	US-08-619-280A-3
10	272	7.8	766	1	US-08-940-391-3
11	257.5	7.4	760	1	US-08-230-491A-2
12	257.5	7.4	760	1	US-08-619-280A-2
13	257.5	7.4	760	2	US-08-940-391-2
14	145.5	4.2	593	5	PCT-US93-07923-11
15	141.5	4.1	588	1	US-07-903-466-3
16	141.5	4.1	588	5	PCT-US93-05794-3
17	129	3.7	305	3	US-08-965-600-1
18	126.5	3.6	317	1	US-07-688-299-1
19	126.5	3.6	317	1	US-07-980-517A-1
20	126.5	3.6	318	1	US-07-688-299-3
21	126.5	3.6	318	1	US-07-688-299-13
22	115.5	3.3	302	3	US-08-965-600-3
23	115	3.3	297	2	US-08-602-359A-37
24	114.5	3.3	579	1	US-08-126-564A-31
25	114.5	3.3	579	5	PCT-US94-09143-31
26	113.5	3.3	900	2	US-08-813-940-4
27	109.5	3.1	528	3	US-08-747-221B-37

28	108.5	3.1	428	1	US-08-190-802A-29	Sequence 29, Appl
29	108	3.1	597	2	US-08-883-534-6	Sequence 6, Appli
30	108	3.1	597	3	US-09-204-764-6	Sequence 6, Appli
31	107	3.1	798	1	US-08-190-802A-64	Sequence 64, Appl
32	107	3.1	798	1	US-08-190-802A-68	Sequence 68, Appl
33	107	3.1	798	2	US-08-308-818-2	Sequence 2, Appli
34	104.5	3.0	626	6	5268290-2	Patent No. 5268290
35	104.5	3.0	1112	2	US-08-714-402-2	Sequence 2, Appli
36	104	3.0	606	2	US-08-883-534-3	Sequence 3, Appli
37	104	3.0	606	3	US-09-204-764-3	Sequence 3, Appli
38	104	3.0	1088	4	US-09-130-242-2	Sequence 2, Appli
39	103.5	3.0	346	2	US-08-602-359A-34	Sequence 34, Appl
40	103.5	3.0	758	1	US-08-258-188-2	Sequence 2, Appli
41	103.5	3.0	758	1	US-08-526-813-2	Sequence 2, Appli
42	103.5	3.0	758	5	PCT-US95-08554-2	Sequence 2, Appli
43	102.5	2.9	428	4	US-09-139-064-2	Sequence 2, Appli
44	102.5	2.9	428	4	US-09-139-064-4	Sequence 4, Appli
45	101.5	2.9	566	3	US-09-264-737-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE REFERENCE: THE SAME
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

Query Match	31.7%	Score 1104.5;	DB 4;	Length 632;
Best Local Similarity	39.9%	Pred. No. 11e-90;		
Matches	255;	Conservative 126;	Mismatches 217;	Indels 41; Gaps 21;
QY	27	RAAYKSVQNEKDSYTSNIWYETKGSVPWTHGKSTDRPSPDGRFLAFIS--DR 84		
DB	25	KAVFQTEISLKDDEYFSKLYLDGKR--VKPFTSGNKDS--NPRFSPNGKLIATSKRDK 81		
QY	85	EGDAAOLYIMSTEGGARKLTIDIPYGVSKPLWSPDGEISILVTISLGEESIDDRKTEQD 144		
DB	82	EGKESELYVIPTDGGEARLLAKFKYGIKNLRFTEGDKSIADVTPFI-----DVEKKGND 134		
QY	145	SVEPVEQGLSYKRGKGLTRGAYAQVLVSVKSGEMKELTSHKADHGDPAFSPDGKWL 204		
DB	135	DVH--IIREIPFWNGVWYKRNVLVDVSGKKRLTPKNLVDOJREFH-NGR-LY 190		
QY	205	FSANLTETDDASKP--HDVYIMSLSGDLKQVTPHRGSGSSFSFSP--DGRYIALLGNEK 260		
DB	191	FTA--QEDREKPLISDLVY--LENKVRKLTp--GKWRILDFLPLDDGFSV-LKANTL 242		
QY	261	EYKNATLSKAWLYDIEQGRITCLITEMLDVHLADALLIGLGAEOPIWTKDSQGFVI 320		
DB	243	ERGIPNAHYHDPDKTKLTKLDNRNAYNSLNSD--VRGSORAEIYKGGWYIVV- 299		

QY 321 GTDGGSTGIYISISGLVPIRLEKEYINSFSLSPDEQHFA--SVTKDPRPSELYSIPL 378
 Db 300 ATDGPRLFRVLDKIERVIGGDSRSEVFDIG-DYIAFTAQDAVT---PTELYIYRD 354
 QY 379 GOEKGOLGANDKFFVREHTISIPERIQATDGVVNVNGLMRPAQMEGETTYPILINING 438
 Db 355 G-KKKVYDFN-KWJGTGYSKPEHFKVKASDGVIEDAWMKPVNFRKGGKYPALILEIHG 412
 QY 439 GPHMMYHTYFHEFOVLAAGYAVVYINPRGSHGYGOEFVNAVRGDYGKDYDDVVMQAVD 498
 Db 413 GKPTAYGYAFMEHHLVLSKGFVIFSNPRGSDGYGEFEAD-IRHGYGRDQDLMEVVD 471
 QY 499 EAIKRDPHIDPKRLGVTGGSGYGFNTNIVGQTNRFKAATVQRTSINWISFPGVSDIGYF 558
 Db 472 EALRFRFIDGERLGVGTGGSGYGFNTNIVGHTNRFKAATVQRTSINWISFPGTDDIGY 531
 QY 559 FTDQLEHDMFEDTEKLDNRSPKLYAANVETPLLIHGERDRCPTEQAEOLFALKKMG 618
 Db 532 FAPDQIGKDPWSNLEGYWEKSPKLYAPNVEPTPLLIHSTEDYRCWLPEALQLFISLKYLG 591
 QY 619 KETLIVRPPNASHNLSRTGHPQRQIRKRLNYISSWFDQHL 657
 Db 592 KRVELAIFPGENHDLRSRSGKPKHRVKRLELIAGWMEKWL 630

RESULT 2

US-08-664-646A-2
 ; Sequence 2, Application US/08664646A
 ; Patent No. 5877001

GENERAL INFORMATION:

; APPLICANT: Murphy et al.
 ; TITLE OF INVENTION: Amidases
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/664,646A
 ; FILING DATE: June 17, 1996

CLASSIFICATION:

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:

FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Charles J. Herron
 ; REGISTRATION NUMBER: 28,019
 ; REFERENCE/DOCKET NUMBER: 331400-53
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-994-1700
 ; TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 622 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; STRANDEDNESS:
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN

US-08-664-646A-2

Query Match

Best Local Similarity 18.2%; Score 635; DB 2; Length 622;

Matches 194; Conservative 113; Mismatches 260; Indels 92; Gaps 27;

QY 18 DPOVAPDGTAAAYVKSOVNOEKDSYTSNIWIYETKTGGSVPWTHGCKRSTDPWSPDGR 77
 Db 19 DPRIR--GNLIATLTITKANMKDKNYESTVVVDELETGS-----RRFIENASMPRI 72
 QY 78 LAFIS--DREGDAALYIMSTEGGGEARKLTDIPYGVSKPLWSPDGESILVTISLGE 136
 Db 73 LAFTCFNEEKETEIIWADIQTLSAKKVLSTK-NVRSQMWNDDSRRLV-----CFKRR 126
 QY 137 DREKTEODSEYEPVEVQGLSKYKRDGKGLTGAYAQVLVYSKSGEMKELTSHKADHGDPA 196
 Db 127 DDEDFVDDVPPVWF-----DNMGFFDGKTTFWLDTAEAEIIE-----QFEKPRF 173
 QY 197 SPQGW----LVTSANLTETDDASKP-----HDVYIMSLESGDLKQVTPHRGSGSSSFS 247
 Db 174 S-SGLHGDALVNVPHRE---GSKPALFKFYDIVLW--KGEKELF-ERVSF--EAVD 224
 QY 248 PDGRYALALGNEKEYKNATLSKAWLYDIEQGRRLCTLE-MLDVHLADALIGD--SLIGA 304
 Db 225 SDGKRILLRGKK--KRFISEHDWLY-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281
 QY 305 EORPIWTKDSOGFVIGTDOGSTGIYISIGLVYPIRLEKEYINSFSLSPDEQHFIA 364
 Db 282 GRVNLWLDGKAERV-----TGDHWT-----YGLDVS DGKALLLI 317
 QY 365 TKDPRSELYSIPLGOEQLTGANDKFRVREHTISIPERIQATDGVVNVNGLMRPAQM 424
 Db 318 MTATRIGELYL--YDGLKQVTEYNGPIFRKLKTFEPHFRFSKD-LEIDGWYLRPEVK 374
 QY 425 EGETTYPILINIHGPHMMYHTYFHEFQVLAAGYAVVYINPRGSHGYGOEFVNAVGRD 484
 Db 375 EEKA--PVIIVFHGGPKGMYCHREVYEMQLMASKYVYVFNPRGSDGYSEDFA 432
 QY 485 YGKDYDDVVMQAVDEATKRDPHIDPKRLGVTGGSGYGFNTNIVGOTNRFKAATVQRTS 544
 Db 433 TGLEDFEDIMNGIEEFKLEPQADRERVGIITGIGYGFMTNNAULTQSLFRAGISENG 492
 QY 545 NWISFHGSDIGYFFTDWQ-----LEHDMFEDTEKLDNRSPKLYAANVETPLLIHGER 598
 Db 493 YLTSYAFSDIGLWY-DVEVIGPNPLENENF---RKL--SPLFYAQNVPKAPILLIHSLE 545
 QY 599 DDCRPIEQAEOLFALKMKGKETHKLVPRPNASHNLSRTGHPQRQIRKRLNYISSWFDQHL 657
 Db 546 DYRCPLDQSLMFYNNVLKDMGKEAYIAIFKRGHGHVSRGSPRHRPKRYRLFIETFERKL 604

RESULT 3

; Sequence 2, Application US/09066285
 ; Patent No. 5985646

GENERAL INFORMATION:

; APPLICANT: Murphy et al.
 ; TITLE OF INVENTION: Amidases
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066,285
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/664,646

FILING DATE: June 17, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Charles J. Herron
 REGISTRATION NUMBER: 28,019
 REFERENCE/DOCKET NUMBER: 331400-53
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 622 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-09-066-285-2

Query Match 18.2%; Score 635; DB 2; Length 622;
 Best Local Similarity 29.4%; Pred. No. 1.6e-48;
 Matches 194; Conservative 113; Mismatches 260; Indels 92; Gaps 27;

QY 18 DPQYADPTAAVYKQVNOEKDSTYNTWIYETKGGVPTWTHGKRSYDPRWSPDGR 77
 Db 19 DPRIR--GNLIAYTLTKANKKONKYEYVVEDLETGS-----RRFTENASMPRISPDGRK 72
 QY 78 LAFIS-DREGDAALYIMSTEGEARKLTDIPYGVSKPLWSPDGESILVTSIGGESID 136
 Db 73 LAFTCFNEEKETEINWADIQTLSAKKVLSTK-NVRSMQWDDSRLLVV-----GFKRR 126
 QY 137 DREKTEQDSYEPVEVOGLSKYKRDGKGLTRGAYAOQLVLSVKSGEMKELTSHKADHGDPAF 196
 Db 127 DDEDFVFDVDDVPWF-----DNMGFFDGEKTFWVLDTEAEIIE-----QFEKPRF 173
 QY 197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLGSDGLKOVTPHRGSGSSSFS 247
 Db 174 S-SGLWHGDAIVVNPVPHRE---GSKPALPKFYDIVLW---KDGEELKF-ERSVF--EAYD 224
 QY 248 PDGRYLALNLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
 Db 225 SDGKRILLRGKKK--KRFISEHDWLW-LWDGELKPIYEGPLDVWEAKLTGKGVYFLTPDA 281
 QY 305 EGRPIWTQDSQGFYVIGTDQSGTGIYIYISIEGLVYPIRLEKEYINSFSLSPDEQHFIAV 364
 Db 282 GRVNLWLDGKAERVV-----TGDHWI-----YGLDVS DGKALLLI 317
 QY 365 TKPDRPELSYPLQGEKQLTGANDKVFYREHTISIPETIYATEDGVVMVNGWLMRPAQM 424
 Db 318 MTATRIGELYL--YDGEKLVQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374
 QY 425 EGETTYPLILNHGPHMMYGHYTFHEFOVLAAGYAVVYINPRSGHGYGOEFVNAVRC 484
 Db 375 EKKA--PVIVFHGPKMGYHRFVEMQMLASKGYVYVFNPRSGDGYSEDFAIRVLER 432
 QY 485 YGKDYDDVMQAVDAIKRPHIDPKRLGVTGGSGYGFMTNMIVGQTNFRKAAVQTSIS 544
 Db 433 TGLEDFEDIMNGIEEFKLEQADREVRGIGTISYGGFMNTNALTQSDLFKAGISENGIS 492
 QY 545 NWISFHGVSIDICYFTDQW-----LEHDMFEDTEKLWDRSLPKYAANVETPLLIHGR 598
 Db 493 YLWTSYAFSDIGLWY-DVEVIGPNLENF---RKL---SPLFYAQNVPKAPILLIHSLE 545
 QY 599 DRCPIEQAEQLFIALKMGKETKLVRFNASHNLRTGHPRIKRLNYISSWFDQHL 657
 Db 546 DYRCPLDQSLMEYNVLKMGKRAYTAIFKRGAGHVSGRSPRHRKRYLRFTEFFERKL 604

RESULT 4

US-09-261-006-2
 ; Sequence 2, Application US/09261006
 ; Patent No. 6004796
 ; GENERAL INFORMATION:
 ; APPLICANT: Murphy et al.

TITLE OF INVENTION: Amidases
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ADDRESSEE: CECCHI, STEWART & OLSTEIN
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/261,006
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/664,646
 FILING DATE: June 17, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Charles J. Herron
 REGISTRATION NUMBER: 28,019
 REFERENCE/DOCKET NUMBER: 331400-53
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 622 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-09-261-006-2

Query Match 18.2%; Score 635; DB 3; Length 622;
 Best Local Similarity 29.4%; Pred. No. 1.6e-48;
 Matches 194; Conservative 113; Mismatches 260; Indels 92; Gaps 27;

QY 18 DPQYADPTAAVYKQVNOEKDSTYNTWIYETKGGVPTWTHGKRSYDPRWSPDGR 77
 Db 19 DPRIR--GNLIAYTLTKANKKONKYEYVVEDLETGS-----RRFTENASMPRISPDGRK 72
 QY 78 LAFIS-DREGDAALYIMSTEGEARKLTDIPYGVSKPLWSPDGESILVTSIGGESID 136
 Db 73 LAFTCFNEEKETEINWADIQTLSAKKVLSTK-NVRSMQWDDSRLLVV-----GFKRR 126
 QY 137 DREKTEQDSYEPVEVOGLSKYKRDGKGLTRGAYAOQLVLSVKSGEMKELTSHKADHGDPAF 196
 Db 127 DDEDFVFDVDDVPWF-----DNMGFFDGEKTFWVLDTEAEIIE-----QFEKPRF 173
 QY 197 SPDGKW----LVFSANLTETDDASKP-----HDVYIMSLGSDGLKOVTPHRGSGSSSFS 247
 Db 174 S-SGLWHGDAIVVNPVPHRE---GSKPALPKFYDIVLW---KDGEELKF-ERSVF--EAYD 224
 QY 248 PDGRYLALNLGNEKEYKNATLSKAWLYDIEQGRLTCLTE-MLDVHLADALIGD--SLIGGA 304
 Db 225 SDGKRILLRGKKK--KRFISEHDWLW-LWDGELKPIYEGPLDVWEAKLTGKGVYFLTPDA 281
 QY 305 EGRPIWTQDSQGFYVIGTDQSGTGIYIYISIEGLVYPIRLEKEYINSFSLSPDEQHFIAV 364
 Db 282 GRVNLWLDGKAERVV-----TGDHWI-----YGLDVS DGKALLLI 317
 QY 365 TKPDRPELSYPLQGEKQLTGANDKVFYREHTISIPETIYATEDGVVMVNGWLMRPAQM 424
 Db 318 MTATRIGELYL--YDGEKLVQVTEYNGPIFRKLKTFEPRHFRFKSKD-LEIDGWYLRPEVK 374
 QY 425 EGETTYPLILNHGPHMMYGHYTFHEFOVLAAGYAVVYINPRSGHGYGOEFVNAVRC 484

Db 375 EEKA--PVIVFHGGPKMGYCHREYVEMQLMASRGYVYVFNPRGSDGYSEDFAIRVLRL 432
QY 485 YGKDYDDVQAVDEAIKROPHIDPKRLGVTGGYGGFMTNWIYVOTNRKAAVOTRSIS 544
Db 433 TGLEDFEDIMNGIEEFKLEPQADREVRGVTIGSYGGFMTNWTQSLDFKAGISENGIS 492
QY 545 NWISFHGVSIDIGYFFTDWQ-----LEHDMPEDETEKLWDRSPLKYAANVETPLILHGER 598
Db 493 YWLTSAFSDIGLWY-DVEVIGPNLENF---RKL---SPLFYAONVKAPILLIHSLE 545
QY 599 DRCPIEQAEQLFALKKMGKTKLVFPNASHNLSTGHPQRKRLNYISSWFDQHL 657
Db 546 DYRCPLDQSLMFYNVLKDMGKEAVIAIFKRGAGHSVRGSPRHRPKRYRLFIEFERKL 604

RESULT 5

US-08-951-088-2
; Sequence 2, Application US/08951088
; Patent No. 6136583
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,088
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/664,646
; FILING DATE: June 17, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-951-088-2

Query Match 18.2%; Score 635; DB 4; Length 622;
Best Local Similarity 29.4%; Pred. No. 1.6e-48;
Matches 194; Conservative 113; Mismatches 260; Indels 92; Gaps 27;
QY 18 DPOYAPDTRAAVYKQVNOEKDSYTNWIVETKGTGGSVPWTHGEKRSIDPRWSPDGRT 77
Db 19 DPRIR--GNLIAYILTKAMKDNKYESTVIVVEDLETGS-----RRFIENASMPRIEFDGRK 72
QY 78 LAFIS-DREGDAAQYIMSTEGEARKLTDIPYGVSKPLWSPDGSILVTISLGEGETSD 136
Db 73 LAFTCFNEEKKEKTEIWAVIDQTLAKVLTSTK-NVRSQWQNDSDRLLV-----GFKRR 126
QY 137 DRETEQDSYEPVEVQGLSYKRDGKGLTRGAYAQVLVSVKSGEMKELTSHKADCDPAF 196

Db 127 DDEDFVDDDDVPWF-----DNMGFFDGEKTFWVLDTEAEIE-----QFEKPRF 173
QY 197 SPDGW---LVFSANLTTDDASKP-----HDVYIMSLGSDGLKQVTPHRGSGSSFS 247
Db 174 S-SGLWGHDAIVVNVPHRE---GSKPALFKFYDIVLW--KDGEEKLF-ERVSE--BAVD 224
QY 248 PDGRYLLALLGNEKEYNATLSKAWLYDIEQRLTCLTE-MLDVHLADALIGD--SLIGGA 304
Db 225 SDGKRILLRGGKK--KRFISEHDLV-LWDGELKPIYEGPLDVWEAKLTEGKVYFLTPDA 281
QY 305 EQRPIWTKDSQSPYVGTGQSGTYIYISIEGLVYIRLEKEYINSLSLSPDEQHFTASV 364
Db 282 GRVNLWLDGKARVY-----TGDHWI-----YGLDVSOGKALLI 317
QY 365 TKPDPSSELYSIPLGOEKKOLTGANDKFVREHTISIPEEIQAYATEDGVNMGWMLRPAQM 424
Db 318 MTATRIGELYL--YDGLKQVTEYNGPIFKLKTFFRHRFRKSKD-LEIDGWILRPEVK 374
QY 425 EGETTYPLIINHGGPHMMYGHYTFHEFQVLAAGYAVVYINPRGSHGYGOEFVNAVGRD 484
Db 375 EEKA--PVIVFHGGPKMGYCHREYVEMQLMASRGYVYVFNPRGSDGYSEDFAIRVLRL 432
QY 485 YGKDYDDVQAVDEAIKROPHIDPKRLGVTGGYGGFMTNWIYVOTNRKAAVOTRSIS 544
Db 433 TGLEDFEDIMNGIEEFKLEPQADREVRGVTIGSYGGFMTNWTQSLDFKAGISENGIS 492
QY 545 NWISFHGVSIDIGYFFTDWQ-----LEHDMPEDETEKLWDRSPLKYAANVETPLILHGER 598
Db 493 YWLTSAFSDIGLWY-DVEVIGPNLENF---RKL---SPLFYAONVKAPILLIHSLE 545
QY 599 DRCPIEQAEQLFALKKMGKTKLVFPNASHNLSTGHPQRKRLNYISSWFDQHL 657
Db 546 DYRCPLDQSLMFYNVLKDMGKEAVIAIFKRGAGHSVRGSPRHRPKRYRLFIEFERKL 604

RESULT 6

PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; FILING DATE: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906


```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-619-280A-3

Query Match 7.8%; Score 272; DB 1; Length 766;
Best Local Similarity 20.4%; Pred. No. 9.3e-16;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

QY 15 SVTDPQYADGTAAAYKVSQVNOEKSYTSNIWIY-----ETKTGSGVPWTHGKRR 65
Db 101 SINDYSISPDGQFILLEYNVVKQWRHSYASYDIYDLNKRQLITEIRIPNNTOWV----- 155
QY 66 STDPKSPDGRTLAFTSDREGDAQAOLYIMSTEGEARKLT-----DIPY----- 109
Db 156 -----TWSVPVGHKLAYVNNND-----IYVKLEPNLPSYRITWTGCKEDILYNGITDWVYEE 206
QY 110 ---GVSKPLWSPGESILVTISLGESEIDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
Db 207 VFSAYSALWSPNGTFLAYA-----QFNDTEVPLIEYSFYSDLSIQ----- 247
QY 165 RGAYAQILVIVSV--KSGEMKELTSHKADHGDPAFSPDGKWLIVFSANLTETDDASKPHDVI 223
Db 248 ---YPKTVRPVYPKAG-----AVNPTVKFFV-----VNTDLSL----- 277
QY 224 MSLESGDLKQVTPHRGSGSSFSFGDGRYLALLGNEKEYKNAT---LSKAWL----- 272
Db 673 LPTPEDNLDHYRNSTVMSRAENFKQVEYLLIHGTADDNVHFOQSAQISKALVGVDFQA 732

```

```

QY 624 VRFPNASHNL-SRTGHPQRKIKRLNVISSWF 653
Db 733 MWYTDDEHGIASTAHQHIIYTHMSHFIKQCF 763

RESULT 10
US-08-940-391-3
; Sequence 3, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-940-391-3

```

```

Query Match 7.8%; Score 272; DB 2; Length 766;
Best Local Similarity 20.4%; Pred. No. 9.3e-16;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

QY 15 SVTDPQYADGTAAAYKVSQVNOEKSYTSNIWIY-----ETKTGSGVPWTHGKRR 65
Db 101 SINDYSISPDGQFILLEYNVVKQWRHSYASYDIYDLNKRQLITEIRIPNNTOWV----- 155
QY 66 STDPKSPDGRTLAFTSDREGDAQAOLYIMSTEGEARKLT-----DIPY----- 109
Db 156 -----TWSVPVGHKLAYVNNND-----IYVKLEPNLPSYRITWTGCKEDILYNGITDWVYEE 206
QY 110 ---GVSKPLWSPGESILVTISLGESEIDRE--KTEQDSYEPVEVQGLSYKRDGKGLT 164
Db 207 VFSAYSALWSPNGTFLAYA-----QFNDTEVPLIEYSFYSDLSIQ----- 247
QY 165 RGAYAQILVIVSV--KSGEMKELTSHKADHGDPAFSPDGKWLIVFSANLTETDDASKPHDVI 223
Db 248 ---YPKTVRPVYPKAG-----AVNPTVKFFV-----VNTDLSL----- 277
QY 224 MSLESGDLKQVTPHRGSGSSFSFGDGRYLALLGNEKEYKNAT---LSKAWL----- 272

```

```

Db 278 -SVTATSTQIT-----APASMLIGDHYLC-----DVTWATQERISLQWLRRIONYSV 324
QY 273 -----YDIEQGRITCLTEMLDHLADALIGDSLIG-ABORPIWTKDSOGFY-VICTDOG 325
Db 325 MDICDYDESSGRWNCILVRAQHIEMSIT-----GWGFRPSEPHFTLDGNSFYKILISNEEG 380
QY 326 STGIYIYISTE--GLVPIRLEKEYINSFSLSPDEQHFIASVTK-----366
Db 381 YRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLYYISNEYKGMPPGGRNLKYIQLSDYT 440
QY 367 -----PDRSELXISPLGQEEK--QL-----TGANDKFRV--EHTIS 399
Db 441 KYTCLSCLEINPER-COYVSFSKCAKYVOLRCSPGLPLYTLHSSVNDKGLRVLEDNSA 499
QY 400 IPEEIYATE-----DGVWVNG--W--LMRPAQMEGETTYPLILNIHGPPHMYGHYF 449
Db 500 LDKMLQNVQMPKSKLDFIILNETKFWYQMLPPHFDKSKYKPLLLDVLVAGPCSQKADTVF 559
QY 450 H---EFQVLAAGYAVVINPRGSHGYGOEFVNAVRYGDKYDDVMOAVDEAIKRDPH 506
Db 560 RLNWATYLASTEINIIVASFDGRGSGYQGDKIMHAINRRILGTFEVBDQTEAARQFSKMG-F 618
QY 507 IDPKRLGVGTGGYGFEMTNWIVGO--TNRKAAVTORSISNWLISFCHVSDIGYFFTDWOLE 565
Db 619 VDNKRIATWGSYGYVYSWLGSGGVFKGCIAPVPSRWEYDVS-----YTERYMG 672
QY 566 HDMFEDTEKLMDRSLPKYAAN--VETPLLIHLGERDDRCPIEQABQLFIALKMKMGKTKL 623
Db 673 LPTPEDNLHDYRNVSTVMSRAENFKQVEYLLIHGTADDNVHFQOQSAQISKALVDVGVDFQA 732
QY 624 VRFPNASHNL-SRTCHPQRKRLNLYISSWF 653
Db 733 MWYDDEHGIIASSTAQHIIYTHMSHIFKOCF 763

```

RESULT 11

```

US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN, AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: FELPE & LYNCH
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; US-08-230-491A-2

```

```

Query Match 7.4%; Score 257.5; DB 1; Length 760;
Best Local Similarity 21.2%; Pred. No. 1.8e-14;
Matches 154; Conservative 119; Mismatches 272; Indels 181; Gaps 32;

```

```

QY 41 SYTSNIWIYETVETGGVSP-----WTHGEKSTDPWSPDGRITLAFISD-----REGDAAQL 91
Db 71 SADNNIVLVNIETGOSYITLSNRTKMSVNASVGLSPDRQFVYLEDSDYSLKWLRYSTATY 130
QY 92 YTMSTEGGEARKLTDIPYGVSKPLSPDCGESILVLTSLGEGSIDDRKTEQDSYEPVEV 151
Db 131 YIYDLSNGEFVGNELPRPIQYLCVSPVSGKLAYVYQ-----NNIYLKQRPDPPFQ-ITF 185
QY 152 QGLSYKRDKGK---GLTRGAYAOVLVSVKSGEMKELTSHKADHGDPF-----FSPDGKWL 203
Db 186 NG---RENKIFNGIPDWVYEEML-----PTKYALWSPNGKFL 221
QY 204 VPSANLTETD-----DASKPHDVVIMSLESGDLKQVTPHRSFGSSSFSPDGRYL 253
Db 222 AY-AEENDKDIPVIAISYVGDQYPRPTINIPYKAGAKNPV-----RIFLIDTTP 272
QY 254 ALLGNEKEYKNATLSK-----AWLYDIEQGRITCLTEMLDVLADAL-----295
Db 273 AVVGQEVVPVAMIASSDYFSWLTVWTDERV-CLOWLKRQNVSVLSICDFREDQWTD 331
QY 296 -----IGDSLIGGA-----EORPIWTKDSQGFYVIGTDQ-GSTGIYIT--SIEGLVYPI 341
Db 332 CPKTOEHIEESRTGWAGGFFVSRPVFSYDAISYYKIFSDKDGKHHYIKDVENAIQIT 391
QY 342 RLEKEYINSFSLSPDE-----295
Db 392 SGKWEAINIFRVTQSLFSYSSNEFEYPPGRNIIYKISGYSPPSKKCVTCHLRKERCQY 451
QY 361 IASVTK-----PDRP-SELY-----SIPLGOEEKOLTGANDKFVREHTISIP- 401
Db 452 TASFSDVAKYVALCYGPGIPFISTLHDGRTDQEIKEENKELENA-----LKNIQLPK 505
QY 402 EEIQAYATEDGVVWVNGWLMRPAQMEGETTYPLILNIHGPHMYGHYF--HEFQVLAAK- 458
Db 506 EEIKKLEVEITLWYKMLIPFPQFDRSKRYPLLIQVYGGQCSQSVRSFAVNWISYLAKSE 565
QY 459 GYAVVYINPRGSHGYGOEFVNAVRYGDKYDDVMOAVDEAIKRDPHDKRLGVGTGS 518
Db 566 GMVIALVDGRGTAFOGDKLLYAVYKLVGYEVEDQITAVRKFIEWG-FIDEKRIATWGS 624
QY 519 YGGFTWNV-IVGQTNRFKAAVQTSISNWLISFHGVSD---IGYFFTDWQLEHDMFEDTEK 574
Db 625 YGGVSSSLALASGTGLFKCGIAPVAVSVSWYVYTERPFGMLPTKDDNLEH--YKNSTV 682
QY 575 LWDRSPLKYAANVETPLLIHLGERDDRCPIEQABQLFIALKMKGKTKLVRFPNASHNLS 634
Db 683 M---ARAEYFENVN---YLLIHGTADDNVHFQOQSAQIAKALVNAQVDFQAMWYSQDNHGLS 737
QY 635 --RTGH 638
Db 738 GLSTNH 743

```

RESULT 12

```

US-08-619-280A-2
; Sequence 2, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESS: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; US-08-619-280A-2

```

```

Query Match 7.4%; Score 257.5; DB 1; Length 760;
Best Local Similarity 21.2%; Pred. No. 1.8e-14;
Matches 154; Conservative 119; Mismatches 272; Indels 181; Gaps 32;

```

```

QY 41 SYTSNIWIYETKGGSPV---WTHGKRSTDPWSPDGRGLAFISD-----REGDAAQL 91
DB 71 SADNNIVLYNIETGQSYTILSNRTMKSVNAGSLSPDRQFVYLEDYSKLWRYSYTATY 130
QY 92 YIMSTEGGEARKITDIPYGVSKPLWSPDGESILVTISLGEGESIDDRKTEQDSYEPVEV 151
DB 131 YYDLSNGEVRGNELPRPQYLCSWSPVSGKLAVVYQ---NNIYLKQRPDGPFPQ-ITF 185
QY 152 QGLSYKRDGK---GLTRGAVQLVLVSKVSGEMKELTSHKADHGDPA-----FSPDGKWL 203
DB 186 NG-----RENKIFNGIPDWVYEEML-----PTKYALMWSNPGKFL 221
QY 204 VFSANLTTED-----DASKPHDVYIMSLGSLKQVTPHRGSGSGSSFPDGRYL 253
DB 222 AY-AEFDKDDIPVATSYGDEQVPTINIPYKAGAKNPV-----RIFIIDTYP 272
QY 254 ALLGNEKEYKNATLSK-----AMLYDIEQGLRLCTLEMLDVHLADAL----- 295
DB 273 AYVCPQEVPPVPMATASSDYFSLWTWTWDERV-CLQWLKRVQNVSVLSICDFREDWOTWD 331
QY 296 -----IGDSLIGGA-----BQPIWTKDSQGFVIGTDQ-GSTGIYI--SIEGLVYPI 341
DB 332 CPKQTEHIEESRTGWAGGFVRFPVFSYDAISYKIFSDKDGKHHIYIKDTVENAIOIT 391
QY 342 RLEKEYINFSLSLSPDE-----QHF 360
DB 392 SGKWEAINIFRVTDQSLFYSNNEFEYVGRNRYISIGSVPPSKKCVTCHLRKRCQY 451
QY 361 IASVTK-----PDRP-SELY-----SIPLOQEEKQLTGANDKFRVREHTISIP- 401
DB 452 TASFSDYAKYVALYCYGPGIPISLHDGRTDQEIKEILEENLENA-----LKNIQLPK 505
QY 402 EEIFQATDGVWNGWLMRPAQMEGETTYPILNTHGPHMMYGHYTF--HEFOVLAAK- 458
DB 506 BEIKKLEVDITLWKMLTPQDFRSKRKYPILLIQVGGPCSQSVRSVFAVNNISYLAKSE 565

```

```

QY 459 GYAVVYINPRGSHGYGOEFVNAVVDYGGKDYDDVMQAVDEALKRPHIDPKRLGVTTGS 518
DB 566 GMVIALVDGRGTAFQDGLLYAVYRKLGVYEVEDQITAVRKFTIEMG-FIDEKRIAIWGS 624
QY 519 YGGEIMTW-IVGQTNRFKAAVTQSRISNNISFHVSD---IGYFFTDWQLEHDMFEDTEK 574
DB 625 YGGYVSSSLALASGTGLFKGIAVAPVSSWEYASVYTERFMGLPTKDDNLEH--YKNSTV 682
QY 575 LWDRSPLKYAANVETPLLLIHGERDDRCPIEQAEQFLALKKMGKETKLVRFNPNASHNLS 634
DB 683 M---ARAERYFRNVD--YLLIHGTADDNVHFQNSAQAKALVNAQVDFQAMWYSDQNHGLS 737
QY 635 --RTGH 638
DB 738 GLSTNH 743

```

```

RESULT 13
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2

```

```

OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-2

```

```

Query Match 7.4%; Score 257.5; DB 2; Length 760;
Best Local Similarity 21.2%; Pred. No. 1.8e-14;
Matches 154; Conservative 119; Mismatches 272; Indels 181; Gaps 32;

```

```

QY 41 SYTSNIWIYETKGGSPV---WTHGKRSTDPWSPDGRGLAFISD-----REGDAAQL 91
DB 71 SADNNIVLYNIETGQSYTILSNRTMKSVNAGSLSPDRQFVYLEDYSKLWRYSYTATY 130
QY 92 YIMSTEGGEARKLTDIPYGVSKPLWSPDGESILVTISLGEGESIDDRKTEQDSYEPVEV 151

```

Db 131 YIYDLSNGEVRGNELPRPIQYLWSPVGSKLAYVYQ-----NNIYLKORPGDPPEQ-ITF 185
QY 152 OGLSYKRDGK-----GLTRGAYAQVLVSVKSGEMKELTSHKADHGDP-----FSPDGKWL 203
Db 186 NG-----RENKIFNGIPDWVEEML-----PTKALWWSNGKFL 221
QY 204 VPSANLTETD-----DASKPHDVYVIMSLEDLQVTPHGRSGSSFSFDPGRYL 253
Db 222 AY-AEFNDKIPVIAYSYVGEQYPTINIPYKAGAKPVV-----RIFIIDTIYP 272
QY 254 ALLGNEKEYKNATLSK-----AWLYDIEQRLTCLTEMLDVLHADAL----- 295
Db 273 AYVGQEVVPAMIASDYYSFWLWTVDERV-CLWLKRVQNVSVLSICDFREDWQWTD 331
QY 296 -----IGDSLIGGA-----EQRPIWTKDSQGFYVIGTDQ-GSTGIYVI--SIEGLVYPI 341
Db 332 CPKQTEHIESRTGWAGGFVSRPVFSYDALSYKIFSDKQYKHIHVIKDTVENAIGIT 391
QY 342 RLEKEYINSFSLSPDE-----OQHF 360
Db 392 SGKWEAINIFRVTQDSLFYSSEFEYPPRRNIYRISIGSYPPSKKCVTCHLKERCOY 451
QY 361 IASVTK-----PDRP-SELY-----SIPLGQEEKOLTCANDKFVREHTISIP- 401
Db 452 TASFSDYAKYVALVYCGPCIDISILHDGRTQOEIKILEENKELENA-----LKNIQLPK 505
QY 402 BEIQAYATEDGVNVLWMLRPAQMEGETTYPLILNIHGPHMMYGHYTP--HEFOVLAAL- 458
Db 506 BEIKKLEVEITLWKYMLPPQFDRSKKYPLLIQVYGGPCSQSVRSFAVWVSYLASKE 565
QY 459 GYAVVYNPRGSHGYGOFFVNAVGRDYGKDYDDVMAQVADRAIKRDHPIDPKRLGVYGG 518
Db 566 GWVLTALVGRGTAQFGDKLLAVYKLGVEVEQITAVRKFIEMG-FIDEKRAIINGWS 624
QY 519 YGGFWTNW-IVGQTRNREKAAVTQRSISNWTISFHGVS-----IGYFFETDOLHEDFDETEK 574
Db 625 YGVVSSILASAGTGLFKCGIAPVSVSEWYASVYTERFMGLPTKDONLEH--YKNSV 682
QY 575 LWDRSPKYAANVTPLILHGERDDRCPIEQABQLFALKMKGETKLVRFPPNASHNLS 634
Db 683 M---ARAEYFRNVD--YLLINGTADNNVHPQNSAQIAKALVNAQVDFQAMWYSDONHGLS 737
QY 635 --RTGH 638
Db 738 GLSINH 743

RESULT 14

PCT-US93-07923-11
; Sequence 11, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US93-07923-11

Query Match 4.2%; Score 145.5; DB 5; Length 593;
Best Local Similarity 19.7%; Pred. No. 0.00014;
Matches 113; Conservative 75; Mismatches 197; Indels 189; Gaps 29;

QY 15 SVTPDPQVAPDCTRAAYKVSQVNOEKDSYTSNIWIY-----ETKGTGGSVPWTHGEKR 65
Db 101 SINDYSISPDQCFILLEYVYKQWRHSYASYDIYDLNKRQLITEERIPNNTQWV----- 155
QY 66 STDPWRSPGDRTLAFISDREGDAAQLYIMSTEGGEARKLT-----DIPY----- 109
Db 156 ---TWSPVGHKLAVWNND-----IYKIEPNLPSYRITWTGKEDIYINGITDWYEE 206
QY 110 ---GVSPWSPDCESILVTISLGEESIDRE--KTEQDSYEPVEVQGLSKYKRGKGLT 164
Db 207 VFSAYSAWLWSPNCTFLAYA-----QFNDTEVPLEIYSFYSDLSQ----- 247
QY 165 RGAYAAQLVLYSV-KSGEMKELTSHKADHGDPFSPDGKWLVSANLTETDDASKPHDYVI 223
Db 248 --YPKTVRPYPKAG-----AVNPTVKFFV-----VNTDSL----- 277
QY 224 MSLESGDLKQVTPHRSFGSSFSFSPDGRYLALLGNEKEYKNAT--LSKAWL----- 272
Db 278 -SVTNATSIQIT-----APASMLIGDHYLC-----DVTWATQERISLWLRRIQYNSV 324
QY 273 -----YDIEOGRCLTEMLDVLADALIGDSLIGG-AEQRPPIWTKDSQGFY-VIGTDQG 325
Db 325 MDICDIDESSGRWNCLVAROHIEMTT-----GWGFRFRPSEPFTLDGNSFYKIISNEEG 380
QY 326 STGIYIYSIE--GLVYPIRLEKEYINSFSLSPDQHFIAVTK----- 366
Db 381 YRHICYFOIDBKDCOTFITKTGWEVIGTEALTSYLYIISNEYKMGPGGRNLYKTLQSDYT 440
QY 367 -----PDRSELYSIPLQEEK--QJ-----TGANDKVR--EHTIS 399
Db 441 KVTCLSCENLPER-CQIYSYVSFSKAKYQLRCSGPGPLPLYTLHSSVNDKGLRVLDSNA 499
QY 400 IPEIOYATE-----DGMVMVG--W-LMRPAQMEGETTYPLILNIHGPHMMYGHYTF 449
Db 500 LDKMLQNVQWPKSKLDFIILNETKFWYOMILPHFDKSKKYPLLLDVIYAGPCOKADTVF 559
QY 450 H---EFQVLAAGYAVVYINPRGSHGYGOEFVNA 480
Db 560 RLNWATVYLASTENTIIVASFDRGSGYQGDKIMHA 593

RESULT 15

US-07-903-466-3
; Sequence 3, Application US/07903466
; Patent No. 5395767
; GENERAL INFORMATION:
; APPLICANT: Murnane, John P.
; APPLICANT: Painter, Robert B.


```
DR InterPro: IPR000379; -.
DR InterPro: IPR001375; -.
DR InterPro: IPR002470; -.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PRINTS: PR00862; PROLIGOPTASE.
FT NON_TER 1
SQ SEQUENCE 663 AA; 76572 MW; 25EF2B47D0CC36DF CRC64;

Query Match 12.6%; Score 439; DB 5; Length 663;
Best Local Similarity 24.5%; Pred. No. 4e-22;
Matches 167; Conservative 119; Mismatches 270; Indels 126; Gaps 28;

QY 42 YTSNIWI-YETKTGSPWTHGKRSRTPRSP-----DGRFLAFISREGDAQAQIYI 93
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 39 YTSQWFEENKT-----KSLAKLTLTETEPIDLVYSEESNFVCLNEG-----VYL 88
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 94 M-----STEGGEARKLTDIPYYSK-PLWSPDGE-SILVTISLGESESIDDRKTEQDS 145
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 89 LHNGQIAKIENDKIITSTPISIDTFEMISVDGKLRGLASMTVPGMTLEESAK----- 143
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 146 YEPVEVGLSKYKRGKGLTGAGAAQV-----LVSVKSGEMKELT-----SH 187
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 144 -KTELEKQNYR-----VYDQLMRWRDWTYDQGFQHLFTFKQEDKSFTEKDIMN 193
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 KADHGDA-----FSPDGKWLVSANLTETDDASKPHDVIYINSLESG-DLKQVT- 235
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 194 DKYDCPAPFGGNEEYDIPDGTGKIAFSI-LLENPASSLDNNVYATFONPLEWNIITI 252
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 PHRGSGSSFSFSPDGR---YIALGNKEKEYKNATLSKAWLYDIBQGRLTCLTEMLD--- 288
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 253 DNKGYNQDLPFNNDGSLLYLSLAPKDESCKSVLKSX---DFKEKIIRDTGKIDLSFS 309
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 289 -----VHLADALIGSLGAE-QRPITWTKDSQGFYVIGTDQGSTGIYIISGLV 338
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 310 APMKIVGNVYHLSQTIEGNYIIVGLDTSKSELTREDVN---IITQKTAGSFVFTTIGII 366
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 339 YPIRLEKEYINSLSDEQHFIASTVTKDRPSELYSIPLGQEKQKLTGANDKVFREHTI 398
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 367 Y-----EY-NSFTL-----PQELF-IFENKVIKQITHINQEVLSITKF 402
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 399 SIPEEQYATDEGVWVNGWLMRPAQMEGETTYPLILNHGGPHMMYGHYTFHEF--QVLA 456
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 403 GEYKEVHYTGANDQIHAFITYPNMTKTKYPVILYTHGGPESPTNFFHYRNQPVIA 462
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 457 AKGYAVVYINPRSGHVGQEFVNAVRYDGYDDVMQAVDEAIKRDPHIDPKRLGVG 516
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 463 AQGYIIVFAPNPHGSGSYGDEFLKAIKRWGWPFDLMKGM DYLTNEPIVDIDNACAMG 522
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 517 GSYGGFMNTNIVGQ--TNRFKAATQRSINWISFHVSDIGYFFTDWQLEHDMFEDTEKL 575
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 523 ASYGGYMMNINSQNTGRFKCIICHGIMDSEGSYYMDMYFL-ETFGYPMYEDNTYY 581
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 576 WDRSPLKYAANVETPLLILHGERDRCPIDQAEQLFTALKMKGKETKLVFRPNASHNLSR 635
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 582 KKYSPLYNVNGYNTPQLTIRHGGSDYRIDVAVGYQQFVTLQKNIESKLVYIPEENHWLVR 641
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 636 TGHPRQRIKRLNISSWFDQHL 657
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 642 ---PYNSIDWHQTVFNWLAKYL 660
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
QY0593
ID QY0593 PRELIMINARY; PRT; 665 AA.
AC QY0593;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DIPEPTIDYL-PEPTIDASE.
GN DPP.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.

OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-1:IMSS.
RA Bruchhaus I., Hellberg A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF059278; AAR20264.1; -.
DR InterPro: IPR000379; -.
DR InterPro: IPR001375; -.
DR InterPro: IPR002470; -.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PRINTS: PR00862; PROLIGOPTASE.
SQ SEQUENCE 665 AA; 76529 MW; D5D71BD45D42AA94 CRC64;

Query Match 12.4%; Score 433.5; DB 5; Length 665;
Best Local Similarity 27.7%; Pred. No. 9.7e-22;
Matches 134; Conservative 84; Mismatches 207; Indels 59; Gaps 16;

QY 197 SPDCKWLVSANLTETDDASKPHDVIYINSLESGDLKQ---VTPHRCGSSGSSFSFSPDGR-- 251
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 215 SPDGTMIASFST-LLENPASSLDNNVYATFEN-PLEWNCITITNNKGYDNQPLFNNDGSL 272
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 -YIALGNKEKEYKNATLSKAWLYDIEQGRLTCLTEMLD-----VHLADALIG 297
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 273 YLSMSAPKDESCKSVLKS---YDFNQKVIDITGNIDLSFAPMKVYVGNVYHLSQTIEG 329
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 298 DSLIGGAP-QRPITWTKDSQGFYVIGTDQGSTGIYIISGLVYPIRLEKEYINSLSLSPD 356
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 330 NVYIVGLDTSKSELTREDVN---IITQKTAGSFVFTTIGIIY-----EY-NSFTL--- 376
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 EQHFIASTVTKDRPSELYSIPLGQEKQKLTGANDKVFREHTIISPEEQYATDEGVWVNG 416
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 377 -----PQELF-LYENKVIKQITHINQEVLSITKFGEYKEIHYTGANDQIHA 422
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 417 WLMRPAQMEGETTYPLILNHGGPHMMYGHYTFHEF--QVLAAGYAVVYINPRSGHSGY 474
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 423 FITYPPNMTKTKYPVILYTHGGPESPTNFFHYRNQPVIAAQYIIVFAPNPHGSGSYG 482
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 475 QEFVNAVRYDGYDDVMQAVDEAIKRDPHIDPKRLGVGSGYGFMTNIVGQ--TNR 533
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 483 DAFLKAIKRWGWPFDLMKGM DYLTNEPIVDIDNACAMGASYGYMMNINSQNTGR 542
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 534 FKAATQRSINWISFHVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKYAANVETPLLI 593
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 543 FKCIICHGIMDSEGSYYMDMYFL-ETFGYPMYEDDTYKYSPLNYVNGYNTPOLT 601
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 594 LHGERDRCPIDQAEQLFTALKMKGKETKLVFRPNASHNLSRTGHPRQRIKRLNISSWF 653
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
DQ 602 IHGSDYRIDVAVGYQQFVTLQKNIESKLVYIPEENHWLVR---PYNSIDWHQVFDWL 658
   |||: | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 654 DOHL 657
   : |
DQ 659 AKYL 662
   : |

RESULT 12
QY9444
ID QY9444 PRELIMINARY; PRT; 591 AA.
AC QY9444;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE S91AA LONG HYPOTHETICAL ACYLAMINO-ACID-RELEASING ENZYME.
GN APE2441.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382956;
```



```

DB 345 KWTGDKTLIVSGEDLGRTRLESF-----PANAKDDYKPKNFTDGGSVSAIYFLPDSS 397
QY 359 HFIA-----VTKDPSELYSIPLGOEEKOLTGANDKFVREHTISIPEEIOYAT 408
DB 398 LLVTSALMTNNVYAKPEKGVIKKIASANEIDPELKLGPDSISEF-----YFO 448
QY 409 EDGVVWNGWLMRPAOMEGETTYPLILNHGGPIMMYGHTYFHEF--QVLAAGYAVVYIN 466
DB 449 GNFTDIHAWIYPENFDKSKYPLIFFIHGGPOGNWADGWSTRNPKAWADOGYVVVAJN 508
QY 467 PRGSHGYGOEFYNAVYGDYKDYDDVMQAVDEAIKRDPHDKRLGVTGGYGGFMTNM 526
DB 509 PTGSGFGQALTTAIONNMGWAPYDLDVKWEYVHENLDYVDTDHGVAAGAGYGGFMINW 568
QY 527 IVGQT--NRFAAVTQRSISNWSHGVSDIGYFFTD-----WOLEHDM-----FE 570
DB 569 ICGSPGLGRKFAL-----VSHDGTFFADAKVSTEEELWFMQREFNGTFWDAR 614
QY 571 DTEKLDWRSPLKAAVNVETPLILHGERDDRCPIEQAEOLFIALKMGRETILVRFPNAS 630
DB 615 DNYRWDPSPERILQFATPMLVIHSDKYRLPVBAGLSLFLNVLQERGVPSRFLNFPDEN 674
QY 631 H 631
DB 675 H 675

RESULT 14
Q9UW98 PRELIMINARY; PRT; 726 AA.
AC Q9UW98 MEDLINE=99009056; PubMed=9792655;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE TRI R 4 ALLERGEN.
OS Trichophyton rubrum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;
OC Trichophyton.
OX NCBI_TaxID=5551;
RN [1]
RP SEQUENCE FROM N.A.
RF MEDLINE=99009056; PubMed=9792655;
RA Woodfolk J.A., Wheatley L.M., Piyasena R.V., Benjamin D.C.,
RA Platts-Mills T.A.E.;
RT "Trichophyton antigens associated with IgE antibodies and delayed type
RT hypersensitivity. Sequence homology to two families of serine
RT proteinases."
RI J. Biol. Chem. 273:29489-29496(1998).
DR EMBL: AF082514; AAD52012.1; -.
DR InterPro: IPR000379; -.
DR SEQUENCE 726 AA; 80120 MW; 83609D55F06783A4 CRC64;

```

```

Query Match 11.5%; Score 402.5; DB 3; Length 726;
Best Local Similarity 22.4%; Pred. No. 1.6e-19;
Matches 169; Conservative 120; Mismatches 245; Indels 221; Gaps 34;

QY 41 SYTSNTIWTETKGGVVPWTHGE-----KRSTDPRWSPDGRTLAFISDREGDAQLYI 93
DB 19 AFTPEDFISAPRGEAIPDPKGLAVFHVSKYNFDKDRPSGNWL--LNLKNGD--INV 73
QY 94 MSTEGGEARKLTDIPGVSKPLWSPGESITLV-----TISLGESESTDDREKTEQDSYEP 148
DB 74 LTWDS-----DVSIITWIGDGTKVYVINGTDSVKGVGWITSD-AKNFGNAYKA 121
QY 149 VEVOG-----LSYKRDGKGLTRG-----AYAOVLVSVK-----177
DB 122 GSVNGAFSGLKLAKSGDKINFGVGSTTKGDLYNEAAAEVSSARIYDSLFRHWDY 181
QY 178 -----SGEMKELTSHK-----ADHGDPAFSPDGKWLVF 205

```

```

DB 182 VGTQFNAVSGTLTKSGDKYSFDGKLKLNLPQVVKYAESPPYPPFGSGDYDLSDGKTAVF 241
QY 206 -----SANTETDDASKPHD-----VYIMSLESGLKOVTPH--RGSGSSSFSPDGR 251
DB 242 MSKAPELKANITTSYIFLVPDHGSRVAEPINKRNGPR---TPQGLEGASSSPVSPDGK 298
QY 252 YLALLGNEXKEYNATLSKAWLYDIEOGRUCLCTEMLDVHLADALIGSLGGAEORPI-- 309
DB 299 RYAYL--QMAKN-----YE-----SDRRVHIAEV-----GTNKPQVR 330
QY 310 -----WTKDSQCFYVIGTDQSGTYIYI-----SIEGLVY 339
DB 331 IASNWDRSPKAVKWSGDRGLTYTAEDHATGKLTLPADARNHKAAPVVKHDSGVSSYF 390
QY 340 PIRLEKEYINSFSLSPDEOHFIASVTKPDPSSELYSIPLGOEEKOLTGANDKFVREHTIS 399
DB 391 IGSSKSVLISGNSLWSNALYOVAT--PDRPN-----RKLFPAN-----EHDPE 431
QY 400 I-----PEETQYATFDG--VMVNGWLMRPAOMEGETTYPLILNHGGPIMMYGHTYFHEF- 452
DB 432 LKGLGPNDIETPLWVDGARTKIHSMIVKPTGFDKNKYVYPLAFLIHGGPQSGWGNWSTRN 491
QY 453 -QVLAAGYAVVYINPRGSHGYGOEFYNAVYGDYKDYDDVMQAVDEAIKRDPHIDPKR 511
DB 492 PRVWADQGYVVAAPNPTGSGTGFGOKLTDDITNDWGGAPYKDLVKIWEHVDHIKIIDTN 551
QY 512 LGVTGGSYGGFMTNWIIVGOT--NRFAAVTQRSISNWSHGVSDIG--YFTDWOLEH 566
DB 552 GIAAGASFGFGVYVNIQOGLGRKFALVSHDG-----TFVGSCKITDELFF-----IEH 602
QY 567 D----MFEDTEKL--WDRSPLKAAVNVETPLILHGERDDRCPIEQAEOLFIALKMGKE 620
DB 603 DFNGTFFEARQNYDRWDCSKPELVAKWSTPQLVYVHNDDFERLSVAGVGLFNVLOEKGV 662
QY 621 TKLVRFPNASHNLSTRGHP---RQRTKRLNVISSW 652
DB 663 SRFLNFPDETHWVTKPENSLVWHQV--LGVWVKW 695

RESULT 15
Q9Y8E3 PRELIMINARY; PRT; 725 AA.
AC Q9Y8E3
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE ALANYL DIPEPTIDYL PEPTIDASE.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RF Doumas A., Monod M.;
RT "Secreted alanyl dipeptidyl peptidase (Dppv) of Aspergillus oryzae."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF125190; AAD41777.1; -.
DR InterPro: IPR000379; -.
DR SEQUENCE 725 AA; 80383 MW; AD5C7FD79F5D666E CRC64;

```

```

Query Match 11.2%; Score 392.5; DB 3; Length 725;
Best Local Similarity 22.1%; Pred. No. 7.7e-19;
Matches 163; Conservative 95; Mismatches 255; Indels 225; Gaps 28;

QY 23 PDGTRAAVYKSOVNOEKSDYTSNIWTETKGGVVPWTHGEKRDSTDPWSPDGRTLAFIS 82
DB 37 PSGDVAFVFSQVSKFKTKTTSQWNVLDLKSG-----DIKLLTND 77
QY 83 D-----REGDAOLYIMSTEGGEARKLTDIPGVSKPLWSPGESI-----123
DB 78 DVSEIWLWGSDDSIIVLVNGTN-----ADIPGVE--LWVSDISDFANGYKAASLPAS 128

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 05:57:53 ; Search time 188.91 Seconds
(without alignments)
6551.237 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaataaagctataaccgc.....catgttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSL1/cgcdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL1/cgcdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL1/cgcdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL1/cgcdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL1/cgcdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL1/cgcdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL1/cgcdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL1/cgcdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL1/cgcdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL1/cgcdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL1/cgcdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL1/cgcdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL1/cgcdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL1/cgcdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL1/cgcdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL1/cgcdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL1/cgcdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL1/cgcdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL1/cgcdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL1/cgcdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL1/cgcdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL1/cgcdata/geneseq/geneseq/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1971	100.0	1971	20	AAX07301
2	227.2	11.5	1896	19	AAV33582
3	149.8	7.6	1869	19	AAV12887
4	49.8	2.5	659	21	AAF13721
5	48	2.4	936	22	AAF58252
6	48	2.4	936	22	AAF58254
7	48	2.4	936	22	AAF58257
8	48	2.4	936	22	AAF58259
9	48	2.4	936	22	AAF58262
10	48	2.4	938	22	AAF58255
11	47.8	2.4	936	22	AAF58252

C 12	47.8	2.4	936	22	AAF58254	Oligonucleotide D1
C 13	47.8	2.4	936	22	AAF58257	Oligonucleotide D1
C 14	47.8	2.4	936	22	AAF58259	Oligonucleotide D2
C 15	47.8	2.4	936	22	AAF58262	Oligonucleotide D2
C 16	47.8	2.4	938	22	AAF58255	Oligonucleotide D1
C 17	39	2.0	6741	21	AAA10595	Gene encoding a su
C 18	37.4	1.9	2196	12	AAQ10958	Gene encoding acyl
C 19	36.2	1.8	9271	19	AAV23080	Fragment HGJ2141 o
C 20	35.8	1.8	1395	18	AAAT97586	Helicobacter pylori
C 21	34.8	1.8	2196	13	AAQ20001	Encodes human acyl
C 22	34.6	1.8	45613	22	AAF28535	Genomic fragment #
C 23	34.2	1.7	729	20	AAZ28861	N.meningitidis nla
C 24	34.2	1.7	10732	21	AAA10594	Gene encoding a su
C 25	33.6	1.7	420	16	AAQ88233	Bovine aFGF synthet
C 26	33.6	1.7	6151	16	AAQ75977	Insulin receptor s
C 27	33.4	1.7	744	21	AAZ53852	Neisseria meningit
C 28	33.4	1.7	957	20	AAZ06344	Cephalosporin acet
C 29	33.4	1.7	2277	19	AAV05370	Human telomerase p
C 30	33	1.7	4108	21	AAAT78099	Human cancer assoc
C 31	33	1.7	5301	18	AAAT93627	Human metastasis-a
C 32	32.4	1.6	906	21	AAAT76786	Human ORFX ORF2341
C 33	32.4	1.6	1232	21	AAZ58188	Arabidopsis phosph
C 34	32.4	1.6	1242	21	AAZ42647	Arabidopsis thalia
C 35	32.4	1.6	2277	19	AAV13834	Homo sapiens ambig
C 36	32.4	1.6	7680	21	AAAX89439	14-3-3 sigma trans
C 37	32.2	1.6	2754	20	AAAX88420	JP11192089 Seq ID
C 38	32.2	1.6	3811	20	AAAX38291	E. coli secA DNA.
C 39	32	1.6	438	20	AAV99905	Fragment of homose
C 40	32	1.6	788	20	AAV99902	Fragment of homose
C 41	32	1.6	1701	21	AAZ49014	Arabidopsis thalia
C 42	32	1.6	2366	10	AAAN1052	Phage DNA insert o
C 43	32	1.6	2381	11	AAQ05241	Sequence encoding a
C 44	31.8	1.6	604	22	AAF68340	Human lung tumour
C 45	31.8	1.6	2046	12	AAQ14249	Cephalosporin acet

ALIGNMENTS

RESULT 1
AAX07301
ID AAX07301 standard; DNA; 1971 BP.
XX
AC AAX07301;
XX
XX 21-MAY-1999 (first entry)
XX
XX Bacillus subtilis serine protease SPL (YuxL) DNA.
DE Serine protease SPL; YuxL; detergent; ss.
XX
XX Bacillus subtilis.
OS
XX
XX W09903984-A2.
XX
XX 28-JAN-1999.
PD
XX
XX 14-JUL-1998; 98WO-US14647.
PF
XX
XX 15-JUL-1997; 97EP-0305232.
PR
XX
XX (GEMV) GENENCOR INT BV.
PA (GEMV) GENENCOR INT INC.
XX
XX Estell DA;
XX
XX WPI; 1999-132231/11.
DR P-PSDB; AAW97789.
XX
XX

Use of serine protease genes from Gram-positive microorganisms - for modification of host cells for the production of heterologous proteins or for producing proteins for use in cleaning compositions

Disclosure; Fig 1A-C; 37pp; English.

This DNA sequence encodes serine protease SP1 (YuxL) of *Bacillus subtilis*. Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see AAW97789-93) were identified via a FASTA search of *Bacillus subtilis* genomic nucleic acid sequences. SP1 was identified by its structural homology to the S9 type serine protease dap2 of yeast. SP2, SP3, SP4 and SP5 were identified by their structural and overall amino acid homology to SP1. Host cells in which the naturally occurring gene encoding one or more of SP1, SP2, SP3, SP4 or SP5 is mutated such that the proteolytic activity is diminished or deleted altogether, can be used for the production of heterologous proteins, e.g. a hormone, enzyme, growth factor, cytokine, protease, carboxylase, lipase, racemase, epimerase, tautomerase, mutase, transferase, kinase or phosphatase (claimed). SP2, SP3, SP4 and SP5 can also be produced on a large scale in a microbial host expression system for use in cleaning compositions such as detergents, bar or liquid soap, dish-care formulations and contact lens cleaning solutions, or for peptide hydrolysis, waste treatment, textile applications, or for peptide fusion-cleavage enzymes in protein production, and as animal feed additives.

Sequence 1971 BP; 558 A; 431 C; 524 G; 458 T; 0 other;

Query Match	100.0%;	Score 1971;	DB 20;	Length 1971;
Local Similarity	100.0%;	Prod. No. 0;		
Matches 1971;	Conservative	0;	Mismatches	0; Indels
1	atgaaaagctgataacgcgacagacacatcacagcgattgtctctgtgaccgatctctcaa	60		
1	atgaaaagctgataacgcgacagacatcacagcgattgtctctgtgaccgatctctcaa	60		
61	tacgccccagacggtacccgtgcgcgatactgaaatccacaagttaaatcaagagaagaat	120		
61	tacgccccagacggtacccgtgcgcgatactgaaatccacaagttaaatcaagagaagaat	120		
121	tcgtatcacatacaatatatggatctatgaaacgaaacggaggagatctgttctctggaca	180		
121	tcgtatcacatacaatatatggatctatgaaacgaaacggaggagatctgttctctggaca	180		
181	catggagaaaagcgaacacgcaccccaagaatggtctccgacggcgacgcttgccttt	240		
181	catggagaaaagcgaacacgcaccccaagaatggtctccgacggcgacgcttgccttt	240		
241	atttctcatcgaagaagcgaatggcgacacactttatatcatcagaagcactaaacacacaa	200		

Qy	661	gtttacataatgtctactcggagttctggagatctttaacaggttacacctcatcgcggtcca	720
Db	661	gtttacataatgtctactcggagttctggagatctttaacaggttacacctcatcgcggtcca	720
Qy	721	ttcggatcaagctcattttccacagacggaagtattcttgcttgccttggaataaataaag	780
Db	721	ttcggatcaagctcattttccacagacggaagtattcttgcttgccttggaataaataaag	780
Qy	781	gaataagaagctacgcctctcaaaaggcgtggtctatgatcgaacaagcgcgcctc	840
Db	781	gaataagaagctacgcctctcaaaaggcgtggtctatgatcgaacaagcgcgcctc	840
Qy	841	acattgttactgagatgctggcagcttcatatttagcggatgcgctgattggagattcatcg	900
Db	841	acattgttactgagatgctggcagcttcatatttagcggatgcgctgattggagattcatcg	900
Qy	901	atcggttggtctgaacagcgcctcatttgacaagaagacacgaagggttttatgtcatc	960
Db	901	atcggttggtctgaacagcgcctcatttgacaagaagacacgaagggttttatgtcatc	960
Qy	961	ggcacagatcaaaagcagtcagggtctctatttatatttctgattgaaggccttggtatccg	1020
Db	961	ggcacagatcaaaagcagtcagggtctctatttatatttctgattgaaggccttggtatccg	1020
Qy	1021	attcgtctggaanaagagtacatcaatagctttctctttccacctgatgaacacacttt	1080
Db	1021	attcgtctggaanaagagtacatcaatagctttctctttccacctgatgaacacacttt	1080
Qy	1081	attgccagtgtagaaagccggacagccagtgtagctttacagtatccccttggaacag	1140
Db	1081	attgccagtgtagaaagccggacagccagtgtagctttacagtatccccttggaacag	1140
Qy	1141	gaagagaacagctgactgcgcggaatgacaagtttgcgaaggagacatacagatatcaata	1200
Db	1141	gaagagaacagctgactgcgcggaatgacaagtttgcgaaggagacatacagatatcaata	1200
Qy	1201	cctgaagagattcaatatgtctacagaagcgcgtgattggtgaacgcctggctgataagg	1260
Db	1201	cctgaagagattcaatatgtctacagaagcgcgtgattggtgaacgcctggctgataagg	1260
Qy	1261	cctgcacaaatgaaagttgagacaacatatccaatttctttaacatacacgcgcgtccg	1320
Db	1261	cctgcacaaatgaaagttgagacaacatatccaatttctttaacatacacgcgcgtccg	1320
Qy	1321	catatgatgcaggacatacatatttcatgagttctcaggttcaggtcgcgcggaagagatac	1380
Db	1321	catatgatgcaggacatacatatttcatgagttctcaggttcaggtcgcgcggaagagatac	1380
Qy	1381	gcggttcgctttatatcaatcccgagagaagccacgcgtcacgcgcggaagtttggatagcg	1440
Db	1381	gcggttcgctttatatcaatcccgagagaagccacgcgtcacgcgcggaagtttggatagcg	1440
Qy	1441	gtcagaggagattatgggggaagatatacagatgtagtcaggcctgtggaagagct	1500
Db	1441	gtcagaggagattatgggggaagatatacagatgtagtcaggcctgtggaagagct	1500
Qy	1501	atcaaacgagatccgcatattgatcctaagcgcgttcggtgtcaacgcgcggaagctacgga	1560
Db	1501	atcaaacgagatccgcatattgatcctaagcgcgttcggtgtcaacgcgcggaagctacgga	1560
Qy	1561	ggctttatgaccaactggatcgtcgggacagacgaaccgctttaagctgcggttaaccag	1620
Db	1561	ggctttatgaccaactggatcgtcgggacagacgaaccgctttaagctgcggttaaccag	1620
Qy	1621	cgtcgatatcaaaattggatcagctttcacgcgcgttcagtgatatcgcgtattctttaca	1680
Db	1621	cgtcgatatcaaaattggatcagctttcacgcgcgttcagtgatatcgcgtattctttaca	1680
Qy	1681	gactggcagcttgacatgacaattgttgagacacagaaaagctctctgggacgcggtctcct	1740
Db	1681	gactggcagcttgacatgacaattgttgagacacagaaaagctctctgggacgcggtctcct	1740

742 WWWW... 683
593 ctgacgcaaatgctgtttctcagctaatattaaactgaacagatgatgccagcagc 652
682 WWWW... 623
653 cgcattgtttacataatgctcagtgatgctgagatcttaagcaggttacacctc 712
622 WWWW... 563
713 gcgctcattcgatcagctcatttcacacgagcggaagtattctgttcttgaa 772
562 WWWW... 503
773 atgaaggaataaagaatgctcagctctcaaggcggtgctctatgatcatcagaac 832
502 WWWW... 443
833 gccgctcacatgtcttactgagatgctgagcttcattttagcggatgctgattgag 892
442 WWWW... 383
893 attcattgatcgtgtgctgacagcgccgatttgacaaagacacagcaggggttt 952
382 WWWW... 323
953 atgtcatcgacagatcaagcagtgacggcatttatttatttgcattgagccttg 1012
322 WWWW... 263
1013 tctaccattcgtctggaagagtgatcatcaatagcttcttcttccactgatgac 1072
262 WWWW... 203
1073 agcacttattgccagtgacaaagccgacagacccgagtgagctttacagtatccgc 1132
202 WWWW... 143
1133 ttggacaggaagaaacagctgactggcgcaatgacaaagtgtgtcaggagacatacga 1192
142 WWWW... 83
1193 tatcaatacctgaagagattcaatgatctacagaa 1227
82 WWWW... 48

RESULT 15
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OB Synthetic.
XX
PN W0200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PR (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;
XX WPI; 2001-159728/16.
DR
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 128; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 2.4%; Score 47.8; DB 22; Length 936;
Best Local Similarity 2.1%; Pred. No. 8.3e-05;
Matches 16; Conservative 396; Mismatches 343; Indels 0; Gaps 0;
QY 473 gggacggcaaaaggctgacgagagtgctgacgcttgcctgacgctgaaagt 532
DB 802 GGG 743
QY 533 cgggtgagatgaagagctgacaaagctgacaaagctgacaaagctgacaaagct 592
DB 742 WWWW... 683
QY 593 ctgacgcaaatgctgtttctcagctaatattaaactgaacagatgatgccagcagc 652
DB 682 WWWW... 623
QY 653 cgcattgtttacataatgctcagtgatgctgagatcttaagcaggttacacctc 712
DB 622 WWWW... 563
QY 713 gcgctcattcgatcagctcatttcacacgagcggaagtattctgttcttgaa 772
DB 562 WWWW... 503
QY 773 atgaaggaataaagaatgctcagctctcaaggcggtgctctatgatcatcagaac 832
DB 502 WWWW... 443
QY 833 gccgctcacatgtcttactgagatgctgagcttcattttagcggatgctgattgag 892
DB 442 WWWW... 383
QY 893 attcattgatcgtgtgctgacagcgccgatttgacaaagacacagcaggggttt 952
DB 382 WWWW... 323
QY 953 atgtcatcgacagatcaagcagtgacggcatttatttatttgcattgagccttg 1012
DB 322 WWWW... 263
QY 1013 tctaccattcgtctggaagagtgatcatcaatagcttcttcttccactgatgac 1072
DB 262 WWWW... 203
QY 1073 agcacttattgccagtgacaaagccgacagacccgagtgagctttacagtatccgc 1132
DB 202 WWWW... 143
QY 1133 ttggacaggaagaaacagctgactggcgcaatgacaaagtgtgtcaggagacatacga 1192
DB 142 WWWW... 83
QY 1193 tatcaatacctgaagagattcaatgatctacagaa 1227

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 02:57:13 ; Search time 2147.34 Seconds
(without alignments)
8676.583 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaagaagctgataacgcg.....catggttgatacaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST :

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

	Query Match	2.4%	Score 46.8;	DB 175;	Length 475;
	Best Local Similarity	52.8%;	Pred. No. 0.0058;		
	Matches 122;	Conservative 0;	Mismatches 108;	Indels 1;	Gaps 1;
OQ	1360	gtgctggcgcgaagattacgcggtcgctttatatcaatccgagagaaggaccagcgctac	1419		
DD	151	GTCTTTGCGAGCAAGGCTACGTCGTGGTGAGCCGACCCAGCAGTACGGGTTAT	210		
OQ	1420	ggcgcaggaatttgtgaa+tgccgttcagaggaagattatggggaaaagattatgacgatgt	1478		
DD	211	GCATGGGCTCTCCAAACCGGNAACTCAAGGCCAGTGGGTGGCAGGCCGTATTGAGCATCT	270		

FEATURES	source
major quality sequence stop: 849.	
Location/Qualifiers	
1. .873	
/organism="Entamoeba histolytica"	
/strain="HM1:IMSS"	
/db_xref="taxon:5759"	
/clone_lib="Entamoeba histolytica Sheared DNA"	
/note="vector: pHS3; Site_1: Bst 1; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."	
327 a	132 c 101 g 313 t
BASE COUNT	
ORIGIN	

1420	ggcaggaatcttgcgaatgcggtcagagagagattatggggaaagattatgacgatatgtg	1479
QY		
872	GGAGATGCATTTTAAAAGCAATTGAAAGAAATGGGAGGATGCCCATTTTGAAGATTTA	813
Db		
1480	atgcagcgtcgtgatgaggctatcaacagagatccgcgatattgatcctcaagcgctcgggt	1539
QY		
812	ATGAAGAGGAATGCATTTTAAAACCTTAGTCAGGCATTAGTTGATATAGATAATGCATGT	753
Db		

The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR08102"

/note="end : TET3"

BASE COUNT 270 a 257 c 203 g 324 t 47 others

ORIGIN

Query Match 1.9%; Score 37.4; DB 219; Length 1101;

Best Local Similarity 59.6%; Pred. No. 6.2;

Matches 59; Conservative 2; Mismatches 38; Indels 0; Gaps 0;

QY 388 ggagaggggaagcattgatgacgagaaacacagcagacacgtatgaacctgtt 447
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| |||||

DB 905 GGGGRAGRAGAAAGGAGGCGGCGGAAAGGAGGAGGAGCAACAAACGAGCAGAT 846
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| |||||

QY 448 gaaatcaagcctctctcaaacggcagcgcaaaagg 486
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| |||||

DB 845 GAGAGGAATGGCTAATTAAGAAATGCAGTGCAGAGGG 807
||||| ||||| || ||||| || ||||| ||||| ||||| ||||| |||||

RESULT 11

AW906913

LOCUS

DEFINITION EST 24-MAY-2000
EST343036 potato stolon, Cornell University Solanum tuberosum cDNA
clone cSTAN22, mRNA sequence.

ACCESSION

AW906913

VERSION

AW906913.1

KEYWORDS

EST.

SOURCE

potato.

ORGANISM

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE

AUTHORS

van der Hoeven, R., Bezzerides, J., Bachem, C., Horvath, B., Visser, R.,

Holt, I.E., Liang, F., Hansen, T.S., Utterback, T., Bowman, C.L., Doan

B., Bougri, O., Buell, C.R., Ronning, C.M., Tanksley, S.D. and Baker

B.

TITLE

Generation of ESTs from potato swelling stolons

UNPUBLISHED (1999)

Contact: Research Genetics, Libraries Division

Tel: 1-800-711-6195

Email: cdna@resgen.com

5 prime sequence.

Location/Qualifiers

1. 513

/organism="Solanum tuberosum"

/cultivar="Bintje"

/db_xref="taxon:4113"

/clone="cSTAN22"

/clone_lib="potato stolon, Cornell University"

/tissue_type="axillary buds of stem explants, swelling

stolons"

/dev_stage="1 to 3 days"

/lab_host="SOLR"

/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:

Xhol; RNA was supplied by Christian Bachem & Beatrix Horvath(Laboratory of Plant Breeding, Dept. of Plant Sciences, Wageningen University, The Netherlands). Total RNA was isolated from developing axillary buds of potato nodal stem cuttings cultured on medium for the introduction of tuber formation as described in Bachem et al. (Plant Journal 1996). Tissue samples were taken of stages corresponding to growing stolons and the early stages of tuber formation."

BASE COUNT 117 a 106 c 122 g 168 t

ORIGIN

Query Match 1.9%; Score 37.2; DB 122; Length 513;

Best Local Similarity 48.2%; Pred. No. 5.4;

Matches 105; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1289 atccactattctaacacacacgagcggtccgcataatgatgacgacacatatcttc 1348
||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||

DB 242 ATCCGCTGATTGATAGTCTTCACGGGGGTCTCATCTGTTTCATTTGTCAGGCTCTCAA 301
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

QY 1349 atgagtttcaggtgctgcgcggaagatacgcggtcgtttatatcaatccgagagaa 1408
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

DB 302 AGCTCTGGCTTTCCTTCTTCACCTGTTATACCTTGTGATTGTAATATAGAGGCT 361
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

QY 1409 gccacggctacggcaggaatttgaatgcggtcagagagattatgggggaaaggatt 1468
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

DB 362 CCTTGGGTTTTGGGGAGGAGCAGTACAAATCTCTCTCGGAAAATTGGATCAGGATG 421
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

QY 1469 atgacatgatgcagcgtgtgagatgagctatcaaa 1506
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

DB 422 TTAATGATGCTGTGCTATAGATCATGTCATTGAA 459
||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

RESULT 12

AZ540694

LOCUS

DEFINITION

ENTDGIOTF

Entamoeba histolytica

Sheared DNA

genomic, DNA sequence.

ACCESSION

AZ540694

VERSION

AZ540694.1

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica.

ORGANISM

Entamoeba histolytica.

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 884)

Loftus, B., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica

HML:IMSS sheared DNA library

UNPUBLISHED (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HML:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 27

High quality sequence stop: 844.

Location/Qualifiers

1. 884

/organism="Entamoeba histolytica"

/strain="HML:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOSI; Site_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

BASE COUNT 315 a 89 c 226 g 254 t
ORIGIN

Query Match 1.9%; Score 37.2; DB 245; Length 884;
Best Local Similarity 50.9%; Pred. No. 6.6;
Matches 115; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

Qy 1364 tggcggcgaaaggtacgcgctgtttatataccacgagaggaagccacggtacgggc 1423
Db 101 TGGCATCAGAGGGTACATTATTATTCACCGCAATAGAGAGGATTACAGGTTTGGAA 160

Qy 1424 aggaatttgtaatgcggcagagagattatgggaaagattatgacatgtatgc 1483
Db 161 TGGAAATGTTAGAGAAATTCATTAGATTATGGAGGCTTTGTATGAAGGATTGTAA 220

Qy 1484 aggcgtggatgaggtatcaaaacgagatccgcattatgacctaaagcggctcggtgca 1543
Db 221 CTTCTGT---TGACATAATGAAGAGAGTATCGTTGATCTGCAGACATTGGATGTG 277

Qy 1544 cggcggaagctacggaggttttatgacaaactgagtcgggca 1589
Db 278 TTGGAGCAAGTTTGGAGGATATTTCGGTTTATTCGCTTGTGGGCA 323

RESULT 13
AZ689494/c
LOCUS ENTIN84TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ689494
VERSION AZ689494.1 GI:11826640
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 926)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 18
High quality sequence stop: 862.

FEATURES
source
1. .926
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

BASE COUNT 266 a 218 c 110 g 332 t
ORIGIN

Query Match 1.9%; Score 37.2; DB 248; Length 926;
Best Local Similarity 50.9%; Pred. No. 6.7;
Matches 115; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

Qy 1364 tggcggcgaaaggtacgcgctgtttatataccacgagaggaagccacggtacgggc 1423
Db 714 TGGCATCAGAGGGTACATTATTATTCACCGCAATAGAGAGGATTACAGGTTTGGAA 655

Qy 1424 aggaatttgtaatgcggcagagagattatgggaaagattatgacatgtatgc 1483
Db 654 TGGAAATGTTAGAGAAATTCATTAGATTATGGAGGCTTTGTATGAAGGATTGTAA 595

Qy 1484 aggcgtggatgaggtatcaaaacgagatccgcattatgacctaaagcggctcggtgca 1543
Db 594 CTTCTGT---TGACATAATGAAGAGAGTATCGTTGATCTGCAGACATTGGATGTG 538

Qy 1544 cggcggaagctacggaggttttatgacaaactgagtcgggca 1589
Db 537 TTGGAGCAAGTTTGGAGGATATTTCGGTTTATTCGCTTGTGGGCA 492

RESULT 14
AZ692448/c
LOCUS ENTH14TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ692448
VERSION AZ692448.1 GI:11829594
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 932)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 454.

FEATURES
source
1. .932
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark,

DB	1301	tcgctcgtgatattctcaaatcctagaggagcgatggtcagggagaggttcgcggat-	1359
QY	1439	cqgtcacaggagagattatgggggaaagattatgacaaatgtgatgcaggctgtgtgagagg	1498
DB	1360	--ataaggccacatctggggagaggattaccaggattcaatgggagggtagtgcgtgaag	1417
QY	1499	ctatcaaacgagatccgcataattgatcctaagcggctcgggtcacgggcgggaagctacg	1558
DB	1418	cattaaaggagatttgaaattcatagatgggaaaggctaggagttacgcgggggttcctatg	1477
QY	1559	gagggttttatgaccaactggatcgtcgggcagacgaacccgctttaaagctgcoggtaccoc	1618
DB	1478	gtggcttcatgacgaactggatagtcggacatacccaacaggttccaaagccgctgtaccoc	1537
QY	1619	agcgcctgatatccaattggatcagctttccagcgcctcagtgatatacggctattctttta	1678
DB	1538	agagatcaattccaatttggataagctcttcctgggacacggatataagttattactttg	1597
QY	1679	cagactggcagcttgagcatgacatggtttgaggacacagaaaagctctgggacgcgtctc	1738
DB	1598	ctccagatcaaataggaaaagatccctggagcaacttggaaaggttatgtgggaaaagacc	1657
QY	1739	ctttaaaaatcgcagcaaaacgtggagacacgcgttttgatactcatggcgagcggagatg	1798
DB	1658	cattaaagtacgtcccaacgttgaaactcccccgtctataatccactcttaccogaagact	1717
QY	1799	accgatgcccgatcagcagcgcgagcagctgtttatcgctctgaaaaaaatgggcaagg	1858
DB	1718	acaggtgttggcttcccgagggtatgcaactcttcataatccctaaaaatcctggggaaga	1777
QY	1859	aaaccaagcttgtcogttttccgaaatgatgcgacaaatttatacgcacgcggacacccaa	1918
DB	1778	gagttgaattggcaattctcccgaggagaaaatcatgacctaaagtatctcgtgggaagcaa	1837
QY	1919	gacagcggtacaaagccctgaattatatacagctcatggttttgatcaacatct	1970
DB	1838	acacacaggtttaaaagacttgaactaatagcagatgatgaggaataatggct	1889

2. RESULT

US-08-664-664A-1
Sequence 1, Application US/08664646A
Patent No. 5877001
GENERAL INFORMATION:
APPLICANT: Murphy et al.
TITLE OF INVENTION: AmigaSes
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAI
ADDRESSEE: CECCHI, STEWART & O
STREET: 6 BECKER FARM ROAD
CITY: ROSLAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/664,
FILING DATE: June 17, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28, 019
REFERENCE/DOCKET NUMBER: 33140
TELECOMMUNICATION INFORMATION:

CORRESPONDENCE ADDRESS:		
ADDRESSEE:	CARELLA, BYRNE, BAIN, GILFILLAN,	
ADDRESSEE:	CECCHI, STEWART & OLSTEIN	
STREET:	6 BECKER FARM ROAD	
CITY:	ROSELAND	
STATE:	NEW JERSEY	
COUNTRY:	USA	
ZIP:	07068	
COMPUTER READABLE FORM:		
MEDIUM TYPE:	3.5 INCH DISKETTE	
COMPUTER:	IBM PS/2	
OPERATING SYSTEM:	MS-DOS	
SOFTWARE:	WORD PERFECT 5.1	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER:	US/09/066,285	
FILING DATE:		

```

RESULT      4
US-09-261-006-1
; Sequence 1, Application US/09261006
; Patent No. 6004796
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILIAN,
; ADDRESSSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

```

```

Query Match      7.6%; Score 149.8; DB 2; Length 1869;
Best Local Similarity 52.3%; Pred. NO. 5.6e-41;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

1291 ccacttattcctaacatacacggcggtccgcatatgatgtacggacatacatatttcat 1350
|| || || || || || || || || || || || || || || || || || || || || ||
1135 CGGTGATAGTCTTCGTCCACGGGGCGGCGAAGGGCATGTACGGACACCGCTTCGTCTAC 1194
|| || || || || || || || || || || || || || || || || || || || || ||

1351 gaggtttcagggtgctggcgcgcaaaaggtacacgcggtcggtttatatcaatccagaggggaagc 1410
|| || || || || || || || || || || || || || || || || || || || || ||
1195 GAGATGCACTGATGGCGGACGACGAAGGGTACTACTGCTCTTCGTCGAACCCGCGCGGCGAGC 1254
|| || || || || || || || || || || || || || || || || || || || || ||

1411 cacggctacggcgaggaattgtgaatcgctcagagagagattatgggggaaaggattat 1470
||||| || || || || || || || || || || || || || || || || || || || || ||
1255 GACGGCTATAGCGAAGACTTCGCGCTCCGCTCTCGGAGAGACTGGCTTCGAGGACTTT 1314
|| || || || || || || || || || || || || || || || || || || || || ||

1471 gacgatgtgatcaggctgtgatgagcctatacaacagagatccgcatattgatccctaag 1530
|| || || || || || || || || || || || || || || || || || || || || ||
1315 GAGACATAATGACAGGCATCGAGGAGTTCCTTAAGTCGAACCGCAGGGCGACAGGGAG 1374
|| || || || || || || || || || || || || || || || || || || || || ||

1531 cggctcgggttcacggcgagctacggaggttttatgaccacactggatcgctcgggcag 1590
|| || || || || || || || || || || || || || || || || || || || || ||
1375 CGCGTTTGAATAACGGGCGATAAGCTACGGCGGCTTCATGACCAACTGGCGCTTGACTGAC 1434
|| || || || || || || || || || || || || || || || || || || || || ||

```

	Query Match	7.6%	Score 149.8;	DB 3;	Length 1869;	
	Best Local Similarity	52.3%;	Pred. No. 5.6e-41;			
	Matches 356;	Conservative 0;	Mismatches 322;	Indels 3;	Gaps 1;	
QY	1291	ccacttattcacaatacacgcggtccgcatatgatctacggacacatacaatat	tttcat	1350		
Dd	1135	CCGGTGATGCTCTGTCTCCACGGCGGCCGAAGGCATGTACGGACACCGCTTCGTCTAC		1194		
QY	1351	gaqtccagtgctgccgcgcaaaagtatacgcgctcggttatataccaacgagagggaagc		1410		
Dd	1195	GAGATCAGCTGATGGCGCAAGGGCTACTTGCCTCTCGTGAACCCGCGCGGCAGC		1254		

APPLICANT: Stephens, David S.
APPLICANT: Kahler, Charlene M.
TITLE OF INVENTION: Neisseria Mutants, Lipooligosaccharides
TITLE OF INVENTION: and Immunogenic Compositions
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/512,955
FILING DATE: 09-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 12-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 729 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
STRAIN: NMB
FEATURE:
NAME/KEY: CDS
LOCATION: 1..729
US-08-512-955-3

Query Match 1.7%; Score 34.2; DB 2; Length 729;
Best Local Similarity 58.3%; Pred. No. 0.19; Mismatches 0; Indels 0; Gaps 0;
Matches 60; Conservative 0

QY 1418 acggcgaggaattgtgaatgcggtcagagagattatgggggaaaggattatgacgatg 1477
DB 206 ACGCCGGGGCGGTGCGGGATGTTTTCCGCGCGGATGATGCGGTATTGAGCCGACGATG 147

QY 1478 tgatcgagggctgtgtagaggctatcaaacagagatccgcata 1520
DB 146 TGTGTGACGAGAGGTTGAGGATTTTGACCCAGACCCCGCGCAT 104

RESULT 9
US-08-557-139-1/c
Sequence 1, Application US/08557139
Patent No. 5827730
GENERAL INFORMATION:
APPLICANT: Pedersen, Oluf
APPLICANT: Bjorbak, Christian
APPLICANT: Frederiksen, Katrine A.
TITLE OF INVENTION: MUTANT DNA ENCODING INSULIN RECEPTOR
TITLE OF INVENTION: SUBSTRATE 1
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827730o No. 5827730disk of No. 5827730th America
STREET: 405 Lexington Avenue

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,139
FILING DATE: 12-FEB-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4041.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6152 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 581...4309
US-08-557-139-1

Query Match 1.7%; Score 33.6; DB 1; Length 6152;
Best Local Similarity 59.4%; Pred. No. 1.5; Mismatches 0; Indels 0; Gaps 0;
Matches 57; Conservative 0

QY 426 gcaggacagctatgaacctgttgaaagtgaaggcctctctacaaacggagcagcaagg 485
DB 914 GTAGGAGAGCCTGGTACCAGCTGCTTGTCTGGGCTCGCTGTCGCCGCGATGCAAAAGT 855

QY 486 gctgacgagaggtgcgtatgccacgctgtgtgtgt 521
DB 854 GCTCGTCCCGGTGTAGAGAGCCACCAGGTGCTTGT 819

RESULT 10
US-08-676-967-2
Sequence 2, Application US/08676967
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676.967
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627

APPLICANT: Stephens, David S.
APPLICANT: Kahler, Charlene M.
TITLE OF INVENTION: Neisseria Mutants, Lipooligosaccharides
TITLE OF INVENTION: and Immunogenic Compositions
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/512,955
FILING DATE: 09-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 12-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 729 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
STRAIN: NMB
FEATURE:
NAME/KEY: CDS
LOCATION: 1..729
US-08-512-955-3

Query Match 1.7%; Score 34.2; DB 2; Length 729;
Best Local Similarity 58.3%; Pred. No. 0.19; Mismatches 0; Indels 0; Gaps 0;
Matches 60; Conservative 0

QY 1418 acggcgaggaattgtgaatgcggtcagagagattatgggggaaaggattatgacgatg 1477
DB 206 ACGCCGGGGCGGTGCGGGATGTTTTCCGCGCGGATGATGCGGTATTGAGCCGACGATG 147

QY 1478 tgatcgagggctgtgtagaggctatcaaacagagatccgcata 1520
DB 146 TGTGTGACGAGAGGTTGAGGATTTTGACCCAGACCCCGCGCAT 104

RESULT 9
US-08-557-139-1/c
Sequence 1, Application US/08557139
Patent No. 5827730
GENERAL INFORMATION:
APPLICANT: Pedersen, Oluf
APPLICANT: Bjorbak, Christian
APPLICANT: Frederiksen, Katrine A.
TITLE OF INVENTION: MUTANT DNA ENCODING INSULIN RECEPTOR
TITLE OF INVENTION: SUBSTRATE 1
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827730o No. 5827730disk of No. 5827730th America
STREET: 405 Lexington Avenue

US-07-688-299-2

Query Match 1.5%; Score 31.8; DB 1; Length 957;
Best Local Similarity 52.7%; Pred. NO. 1.5;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1489 gtggtatgagcgtatcaaacagagatccgcatattgatcctaagcgctcggtgtcacgggc 1548
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 GCGGCTTCAGGTCATCAGCAGCTTCGACGAGTTGACGAACAAGATCGGTGTGACAGGA 537
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1549 ggaagctacgaggtttttatgaccaactggatcgtcgggcagacgaacccgttttaaagct 1608
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 538 GGAAGCCAAAGCGGAGGTTTAACCATTTGCCGCAGCAGCGGCTGTGACAGACATTCCAAAGGCC 597
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1609 gcggtaccaca 1619
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 598 GCGGTTGCCGA 608
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-07-980-517A-2
; Sequence 2, Application US/07980517A
; Patent No. 5338676
; GENERAL INFORMATION:
; APPLICANT: Kenji, MITSUSHIMA
; APPLICANT: Akio TAKIMOTO
; APPLICANT: Shigeo YAGI
; APPLICANT: Takayasu SONOYAMA
; TITLE OF INVENTION: Cephalosporin Acetylhydrolase Gene
; TITLE OF INVENTION: and Protein Encoded by Said Gene
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5338676th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07980,517A
; FILING DATE: 19921123
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 113483/1990
; FILING DATE: 27-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gerald M. Murphy, Jr.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-3177
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 957 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-07-980-517A-2

478 GCGCTTAGGTCATCAGCAGCTTCGACGAGGTTCGACGAGGATCGGTGTGACAGGA 537
1549 ggaagctacggaggttttatgaccaactggatcgcggagacgaaccgcttttaagct 1608
538 GGAAGCCCAAGCGGAGGTTTAACCATTCGCCGACGAGCGCTGTGACACATTTCCAAAGCC 597
1609 gccgttaccga 1619
598 GCGTTGCCGA 608

Qy 1609 gccgttaccga 1619
Db 781 GCGTTGCCGA 791

Search completed: September 26, 2001, 08:00:14
Job time: 17436 sec

RESULT 15
US-07-688-299-12
; Sequence 12, Application US/07688299
; Patent No. 5281525
; GENERAL INFORMATION:
; APPLICANT: MITSUSHIMA, Kenji
; APPLICANT: TAKIMOTO, Akio
; APPLICANT: YAGI, Shigeo
; APPLICANT: SOMOYAMA, Takayasu
; TITLE OF INVENTION: Cephalosporin Acetylhydrolase Gene and
; TITLE OF INVENTION: Protein Encoded by Said Gene
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 301 N. Washington Street, P. O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,299
; FILING DATE: 19910422
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-113483
; FILING DATE: 27-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY, JR., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-2765P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2046 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 184...1140
US-07-688-299-12

Query Match 1.6%; Score 31.8; DB 1; Length 2046;
Best Local Similarity 52.7%; Pred. No. 2.7;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Qy 1489 gtggatgaggtatcaaacagagatccgcataattgatcctaaagcgctcgggtcagggc 1548
Db 661 GCGCTTAGGTCATCAGCAGCTTCGACGAGGTTGACGAAACAAGGATCGGTGTGACAGGA 720
Qy 1549 ggaagctacggaggttttatgaccaactggatcgcggagacgaaccgcttttaagct 1608
Db 721 GGAAGCCCAAGCGGAGGTTTAACCATTCGCCGACGAGCGCTGTGACACATTTCCAAAGCC 780

us-09-462-845-1.rni

Wed Sep 26 11:29:25 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:34:53 : Search time 34.68 Seconds
(without alignments)
1148.499 Million cell updates/sec

Title: US-09-462-845-2

Perfect score: 3489

Sequence: 1 MKKLITADDTAIVSVTDPO.....HPRQIKRLNVISSWFDQHL 657

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:
	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3489	100.0	657	20	AA08553
2	3489	100.0	657	20	AAW97789
3	1104.5	31.7	632	19	AAW70508
4	625	17.9	622	19	AAW41248
5	590.5	16.9	684	21	AAAB18512
6	453	13.0	841	21	AAAB18511
7	396.5	11.4	783	10	AAAP94265
8	391	11.2	732	13	AAAR11056
9	383	11.0	732	13	AAAR20001
10	311	8.9	541	21	AAAB18510
11	272	7.8	759	15	AAAS4612

12	272	7.8	766	14	AAAR0909	Sequence encoded b
13	271	7.8	766	15	AAAS4611	Native CD26. Homo
14	269	7.7	766	15	AAW89596	Saccharomyces cere
15	262.5	7.5	732	21	AAAB18507	Amino acid sequenc
16	257.5	7.4	760	18	AAW27438	Human fibroblast a
17	247.5	7.1	739	15	AAAS4613	Delta24-34 CD26.
18	244.5	7.0	767	21	AAAB1748	Rat dipeptidyl pep
19	243	7.0	771	20	AAW89589	Aspergillus oryzae
20	238	6.8	771	20	AAW97798	Aspergillus oryzae
21	237.5	6.8	771	20	AAW89589	Amino acid sequenc
22	235	6.7	731	21	AAAB18515	Arabidopsis thalia
23	232.5	6.7	731	21	AAAG38773	Arabidopsis thalia
24	231.5	6.6	621	21	AAAG28040	Arabidopsis thalia
25	223.5	6.4	569	21	AAAG38775	Arabidopsis thalia
26	222.5	6.4	569	21	AAAG28041	Arabidopsis thalia
27	219	6.3	759	18	AAW31963	Human fibroblast a
28	212.5	6.1	508	21	AAAB42928	Human ORFX ORF2692
29	201.5	5.8	960	21	AAAG48128	Arabidopsis thalia
30	201.5	5.8	968	21	AAAG48127	Arabidopsis thalia
31	201.5	5.8	1063	21	AAAG48126	Arabidopsis thalia
32	191.5	5.5	487	21	AAAG28042	Arabidopsis thalia
33	184.5	5.3	710	21	AAAB29603	Arabidopsis thalia
34	182.5	5.2	710	21	AAAB29602	Human prolyl endop
35	179.5	5.1	652	21	AAAG29115	Porcine prolyl end
36	179.5	5.1	657	21	AAAG29114	Arabidopsis thalia
37	166.5	4.8	615	21	AAAB29604	Pyrococcus furiosu
38	164.5	4.7	690	21	AAAB29599	Aeromonas punctatu
39	162.5	4.7	723	19	AAW40286	S. capsulata proly
40	160.5	4.6	668	22	AAAB79081	Corynebacterium gl
41	160.5	4.6	690	21	AAAB29600	Aeromonas hydrophi
42	156.5	4.5	351	21	AAAG48134	Arabidopsis thalia
43	156.5	4.5	360	21	AAAG48133	Arabidopsis thalia
44	156.5	4.5	381	21	AAAG48132	Arabidopsis thalia
45	154	4.4	295	21	AAAG48135	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA08553
ID AA08553 standard; protein; 657 AA.
XX AC
XX AA08553;

DT 03-AUG-1999 (first entry)
XX

DE B. subtilis hydrolase protein YUXL.
XX

KW Hydrolase; bacterial; Gram positive; YUXL; YUMA; YIVV; TOKD; YCLE; YTPA; YTPA;
KW YDYN; YBPK; YFHM; YDJP; YVFP; YVAM; YQUL; SRFAD; YCGS; YBAC; YUII;
KW YODD; YJCH; YODH; detection; cleaning; detergent; textile treatment;
KW animal feed; peptidase hydrolysis; waste treatment; cleavage.
XX
OS Bacillus subtilis.
XX
PN WO9927081-A2.
XX
PD 03-JUN-1999.
XX
PF 19-NOV-1998; 98WO-US24973.
XX
PR 20-NOV-1997; 97GB-0024629.
XX
PA (GENV) GENECOR INT INC.
XX
PI Estell DA;
XX
DR WPI; 1999-347714/29.
XX
XX Hydrolases from Gram-positive microorganisms
XX
XX Claim 1; Page 20; 24pp; English.

QY 301 IGGAEQRPWTKDSQGVYVIGTDQSTGIYYSIEGLVPIRLEKEYINSFSLSPDEQHF 360
 DQ 301 IGGAEQRPWTKDSQGVYVIGTDQSTGIYYSIEGLVPIRLEKEYINSFSLSPDEQHF 360
 QY 361 IASVTKPDRPSELYSIPLGQEQKLTGANDKPVREHTTISPEETQYATPDGVMVNGWLMR 420
 DQ 361 IASVTKPDRPSELYSIPLGQEQKLTGANDKPVREHTTISPEETQYATPDGVMVNGWLMR 420
 QY 421 PAQMEGETTYPILINIGGPHMYGHTYFHEFOVLAAGYAVVYNPRGSHYGGEFVNA 480
 DQ 421 PAQMEGETTYPILINIGGPHMYGHTYFHEFOVLAAGYAVVYNPRGSHYGGEFVNA 480
 QY 481 VRDYGKDDVDMQAVDEAIKRDHPIDPKRLGVTGGSYGFMFNWVGQTNRFKAATQ 540
 DQ 481 VRDYGKDDVDMQAVDEAIKRDHPIDPKRLGVTGGSYGFMFNWVGQTNRFKAATQ 540
 QY 541 RSTSNWISFHVGSIDIGYFFTDWLEHDMFEDTEKLWDRSLPKYAANVETPLLIHGERDD 600
 DQ 541 RSTSNWISFHVGSIDIGYFFTDWLEHDMFEDTEKLWDRSLPKYAANVETPLLIHGERDD 600
 QY 601 RCPTEQAQELFIALKMGKETKLVRFNASHNLSRTGHPQRIRKRLNYISSWFDQHL 657
 DQ 601 RCPTEQAQELFIALKMGKETKLVRFNASHNLSRTGHPQRIRKRLNYISSWFDQHL 657

RESULT 3

AAW70508
 ID AAW70508 standard; Protein; 632 AA.

XX AAW70508;

DT 29-DEC-1998 (first entry)

DE Pyrococcus horikoshi acylpeptide hydrolase.

XX Pyrococcus horikoshi acylpeptide hydrolase.

OS Pyrococcus horikoshi.

PN JP10210977-A.

PD 11-AUG-1998.

PE 31-JAN-1997; 97JP-0018381.

PR 31-JAN-1997; 97JP-0018381.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX WPI; 1998-488369/42.

DR N-PSDB; AAV33582.

PT A heat-resistant acylpeptide hydrolase and a gene coding it - useful for hydrolysing the C termini of proteins at high temperature

XX Claim 2; Pages 5-6; 8pp; Japanese.

XX The present sequence represents a Pyrococcus horikoshi acylpeptide hydrolase. The enzyme has the following properties: (a) it hydrolyses an acylpeptide; (b) an optimum temperature of 90-95 degrees Centigrade; (c) an optimum pH of 5.0-6.0; (d) no loss of activity when heated to 95 degrees Centigrade for 3 hours at pH 7.5; and (e) a molecular weight of 60 kDa. The enzyme is claimed useful for hydrolysing the amino end of an acylated protein and a peptide at high temperatures.

XX Sequence 632 AA;

Query Match

Best Local Similarity 31.7%; Score 1104.5; DB 19; Length 632;

Matches 255; Conservative 126; Mismatches 217; Indels 41; Gaps 21;

QY 27 RAAVYKSOVNORKSDYTSNIWIYETKGGSPWTHGCKRSTDPWSPDGRITLAFIS--DR 84
 DQ 25 kavfqtelslkdddyfsklylydgkr--vkpftsgnkds-nprfsgnkllafstskrdk 81
 QY 85 EGDAAOALVIMSTEGEARKLTDIPYGVSKPLMSPOGESILVTISLGESESIDDRKTEQD 144
 DQ 82 egkeselyvptdggearllakfygiknrlftedgkslavvtpl-----dvekkgnp 134
 QY 145 SYEPVEVQGLSVKRDGKGLTGAYAOALVSVKSGEMKELTSHKADHGDPAPSPDGKWL 204
 DQ 135 dvh--ilreipfwngvwygkrnvvylvdesgkkrktkpnlnvdqirfrh-ngr-ly 190
 QY 205 PSANLTETDDASKP--HDVYIMSJESDLKQVTPHRGSGSSSSP--DGRYLALLGNEK 260
 DQ 191 fta---qedrerkpilsdlyv--lenrkrvkltp--gkwrildflplddgsfv-ikantl 242
 QY 261 EYKNATLSKAWLYDTEOGRLTCLTEMLDVLADALIGSLGGAEORPIWTKDSOGFYVI 320
 DQ 243 ergiptnahiyhydpktgelkkltkddrnaynslnsd--vrgsraelvkyegwiyv- 299
 QY 321 GTDOGSTGIYVISIEGLVPIRLEKEYINSFSLSPDEQHFIA--SVTKPDRPSELYSIP 378
 DQ 300 atdpranlfrnldgkierviggrsvesfdig-dylafatqadv----ptelyiyrd 354
 QY 379 GOEKEQLTGANDKPVREHTTISPEIQYATPDGVMVNGWLMRPAQMEGETTYPILINIG 438
 DQ 355 g-kekvtvdfn-kwikgtiskpvhkvdgveidawmkpvnfrkgykypallelbg 412
 QY 439 GPHMYGHTYFHEFOVLAAGYAVVYNPRGSHYGGEFVNAVRGDDYDVMQAVD 498
 DQ 413 gpklaygafmhefvltskfvvfnprgsgdygeefad-irghygerdyqdlmevvd 471
 QY 499 EAIKRDHPIDPKRLGVTGGSYGFMFNWVGQTNRFKAATQRTSISNWSFHVGSIDIGY 558
 DQ 472 ealrrfoidgerlgtvgsgygmfnwlvghnrfkaavtqrsisnwsifsgttdigy 531
 QY 559 FTDWLEHDMFEDTEKLWDRSLPKYAANVETPLLIHGERDDRCPIEAOQLFIALKMG 618
 DQ 532 fapdqigkdpwnlegwyekspkypnvetplliihstedyrcwlpaelqflslkylg 591
 QY 619 KETKLVRFNASHNLSRTGHPQRIRKRLNYISSWFDQHL 657
 DQ 592 krvelaifpghndlsrsgkphkrvrkrllelaqwmekwl 630

RESULT 4

AAW41248
 ID AAW41248 standard; Protein; 622 AA.

XX AAW41248;

DT 22-JUN-1998 (first entry)

DE Thermococcus amidae.

XX Amidase; thermostable enzyme; optically active compound; L-amino acid; peptide; peptidomimetic; archaeobacterium.

OS Thermococcus sp. strain GU5L5.

XX Key Location/Qualifiers

FT Misc-difference 410

FT /note= "encoded by TGC"

FT Misc-difference 411

FT /note= "encoded by TGC"

PN W09748794-A1.

XX 24-DEC-1997.

XX 17-JUN-1997; 97WO-US09319.

XX 17-JUN-1996; 96US-0664646.

XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PA Murphy D, Reid JC, Robertson D;
XX PE WPI; 1998-063135/06.
XX DR N-PSDB; AAV12887.
XX PA Nucleic acid encoding heat-stable amidase from Thermococcus GU515 -
XX PA useful for removing N-terminal amino acids from synthetic peptide(s)
XX PA and peptidomimetic(s)
XX PS Claim 1; Page 39-41; 56pp; English.
XX CC This protein comprises a novel thermostable amidase of the
XX CC archaeobacterium Thermococcus GU515. Its amino acid sequence was
XX CC deduced from an isolated nucleic acid (see AAV12887). Recombinant
XX CC amidase can be produced in transformed host cells, especially
XX CC prokaryotic cells. The amidase is used to remove Arg, Phe or Met
XX CC from the N-terminus of synthetic peptides or peptidomimetics
XX CC (claimed). Removal of the N-terminal residue can be done even in
XX CC the presence of a more reactive ester bond (very difficult to
XX CC achieve non-enzymatically). The amidase is selective for L-amino
XX CC acids and can therefore be used to produce optically active
XX CC compounds. The protein tolerates temperatures up to at least 70
XX CC degC and high concentrations of organic solvent (e.g. over 40%
XX CC dimethyl sulphoxide) so can cleave bonds that are normally
XX CC resistant.
XX PS Sequence 622 AA;
XX SQ
Query Match 17.9%; Score 625; DB 19; Length 622;
Best Local Similarity 29.1%; Pred. No. 3.9e-44;
Matches 192; Conservative 113; Mismatches 262; Indels 92; Gaps 27;
QY 18 DPQYAPGCTRAAYKVSQVNEKSYTNIWYETKTCGVSVPWTHGKRSKSTDPKSPDGRT 77
DB 19 dprir--gnliaytltkannkndkystvvdletys-----rrfienasmpriispdgrk 72
QY 78 LAFIS-DREGDAAQLYIMSTEGGARKLTDIPYGVSKPLWSPDGESILVITSLGEGESID 136
DB 73 laftcfeeketeiawdiqtltsakvstsk-nvrsmqwnddrrllv-----gfkrr 126
QY 137 DRETEQDSPEPVEVQGLSKYKRCGKGLTRGAYAQVLVSVKSGEMKELTSHKADHGDPAP 196
DB 127 ddeidvfdvdpvfvf-----dmgffdgkttfwdtdaeaiie-----qfexprf 173
QY 197 SPDGKW-----LVFSANLTETDDASKP-----HDVYIMSLSGDLKQVTPHPRGSGSSSFS 247
DB 174 s-sglwhgdaivnnvphre---gskpalfkfydivlw--kdgeeklf-ervsf--eavd 224
QY 248 PDGRYLALLGNEKEYKNATLSKAWLYDIEQGLTCLTE-MLDVHLADALIGD--SLIGGA 304
DB 225 sdgkrillrgkkk--krfishndwly-lwdgelkpiyegpldwweakitegkvylftpda 281
QY 305 EQRPLTWKDSQGVYVIGTDOGSTGIYIYISIEGLVYPIRLEKEYINFSLSLSPDEQHFIAV 364
DB 282 grvnllwldvgkaervv-----tgdhwi-----yglvdsdgkallli 317
QY 365 TKPDRPELSYIPLGOEKQLTGANDKVFREHTISIPREIQVATEDGVNNGWLMRPAQM 424
DB 318 mtratrigeiyl--ydgelkqvteyngpifrkltktrfhrfkskd-leidgwyirpevk 374
QY 425 EGETYPPLINTHGGPHMYGHTYFHEFOVLAAGYAVVYINPRSGHGYGQEFVNAVGRD 484
DB 375 eeka--pvlvfvngpkmghrfvymqlmaskyyccfvnprgsdysedfalrvler 432
QY 485 YGCKYDDVQVQVDAIKRPHIDPKRLGVGTGSGYGGFMNIVGQTNRFKAATVQRTSIS 544
DB 433 tgldefedlmngieeffklepdaqdervgitgisyggfntnwtaltqtdlftkagisengis 492
QY 545 NMISFHGVSIDIGYFTDQW-----LEHDMFEDTEKLDWRSPLKYAANVETPLLILHGER 598

Db 493 ywltsyafsdiglw-dveigvnpnenf---rkl---splfaqnvkapillihsie 545
QY 599 DDRCPIEQAEQLFIALKMGKETKLVRFNASHNLSRTGHPRQRIKRLNVISSWFDQHL 657
Db 546 dyrcpldqslmfynvlkdmgkeayiaifkrgahghsvrgsprhrkryrlfiefkerkl 604
RESULT 5
AAB18512
ID AAB18512 standard; protein; 684 AA.
XX AC AAB18512;
XX DT 15-JAN-2001 (first entry)
XX DE H3 homologue of prolyl-tripeptidyl peptidase DPP.
XX KW Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
XX KW gingivitis; periodontitis.
XX OS Porphyromonas gingivalis.
XX PN WO200052147-A2.
XX PD 08-SEP-2000.
XX PF 03-MAR-2000; 2000WO-US055551.
XX PR 05-MAR-1999; 99US-0123148.
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX PA (TRAV/) TRAVIS J.
XX PA (POTE/) POTEMPA J.
XX PA (BANB/) BANBULA A.
XX PI Travis J, Potempa J, Banbula A;
XX WPI; 2000-594181/56.
XX PT Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
XX PT for identifying its inhibitor which is useful for protecting an animal
XX PT from a periodontal disease such as gingivitis and periodontitis -
XX PS Claim 23; Fig 6; 58pp; English.
XX CC The present sequence represents a H3 homologue of a prolyl
XX CC tripeptidyl-peptidase (designated DPP) from Porphyromonas gingivalis.
XX CC The prolyl tripeptidyl-peptidase has an amidolytic activity, and cleaves
XX CC a peptide bond in a target polypeptide having at least 4 amino acids.
XX CC This bond is between a proline and an amino acid attached to the
XX CC alpha-carboxyl group end of the proline. The polypeptide is useful
XX CC for identifying inhibitors. These inhibitors are then useful for
XX CC reducing the growth of bacterium or for protecting an animal from a
XX CC periodontal disease such as gingivitis and periodontitis caused by
XX CC Porphyromonas gingivalis.
XX SQ Sequence 684 AA;
Query Match 16.9%; Score 590.5; DB 21; Length 684;
Best Local Similarity 25.8%; Pred. No. 3.8e-41;
Matches 179; Conservative 135; Mismatches 298; Indels 81; Gaps 23;
QY 5 ITADDTAIVSYTDQYAPDGTAAVYKVSQVNEKDSYTSNIWYETKTCGVSVPWTHGK 64
DB 31 ltpelfmtlrsrvsealspdgtkavysfvdvktkatreiftvnlsgsrkqitdtes 90
QY 65 RSTDPRWSPDGRTLAFISDREGDAAQLYIMSTEGGARKLTDIPYGVSKPLWSPDGESIL 124
DB 91 neyapawmadgkriafmsn-eggsmqlwmmnadtgrrqlsnleggitflfspddekvl 149
QY 125 VTISLGEGESIDREKTEQDSYEPVE-----VQGLSYKR-----DGK 161

Db 150 fckdikfg-----krtdkdiypdikatkgrittdlmykhwedwvvetiphipfianatdgm 202
 QY 162 GLTRGAYAOQLVSVKSGEMKELTSHKADHG--DPAFSPDGRKWLVSANLTETDDA--S 216
 Db 203 -itgk-----dimegepye-ampkpwsgiedfswpgdgnaiay-asrkkgtmaysls 252
 QY 217 KPHDYVIMSLSGDLKQVTPHRGSGSS--FSPDGRYALLGNEKEYKNATLSKAWLY 273
 Db 253 tnsdiynlasgrthnis--egmmgydytkfspdgskslawismerdgyesdlkrlfva 310
 QY 274 DIEOGRUCLTEMLDVHLADALIGSLGGAQRPIWTKDSQGFYVIGTDCSTGIYVIS 333
 Db 311 dlatgkrthvntfdynv-----dmi-----qwapdskgyiflackeaetnlwit 356
 QY 334 IE-GLVYPIRLEKEYINSFSLSPDEOHFIASVTPDRPSELYSIPLGQEEKQLTGANDKF 392
 Db 357 lktgkirtqgqhdyaadfsvrndv--mlakrhsefelpddlyrnlknagaagavtaenk 414
 QY 393 VREHTISIEPIQY-ATEDGVYVNGWLMRPAQMEGETTYPILNTHGGPHMMYGH--TYF 449
 Db 415 ildrltptlctekrmktdtgnmltwvlpnfdknkypailycqgqgntvsgfwsfr 474
 QY 450 HEFOVLAAGYAVVYINPRGSHGCGEYFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDP 509
 Db 475 wnlrmaeqyviapnhrhvgpfgqkneqisdygqnmrdyiltavde-mkkepyvvg 533
 QY 510 KRIGVTGSGYGGFMNIVG--QTNRFKAATVQORSINWISFHGVSDDIGYFTDQLEHDM 568
 Db 534 drigavasyggsfvywlaghdkrfaafiahagifn-lemgyattcemfawndigpff 592
 QY 569 FED----TEKLWDSPLKYAANVETPLLIHGERDRRCPIEQAPQLFKKMGKETKLV 624
 Db 593 wekdnnvagtatsphkfgnwdtpilmlhgeldfilasqamaafdaaqlgvpsml 652
 QY 625 RFPNASHNLSRTGHPRIKRLNYSISSWFOHL 657
 Db 653 iypdenhwlq---pqnalifhrtfgwldrw 682

RESULT 6

AA18511 standard; protein; 841 AA.
 AC AA18511;
 DT 15-JAN-2001 (first entry)
 DE H2 homologue of prolyl-tripeptidyl peptidase DPP.
 KW Prolyl tripeptidyl-peptidase; amidolytic activity; periodontal disease;
 XX gingivitis; periodontitis.
 OS Porphyromonas gingivalis.
 XX WO200052147-A2.
 PN 08-SEP-2000.
 PD 03-MAR-2000; 2000WO-US05551.
 PF 05-MAR-1999; 99US-0123148.
 PR (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA (TRAV/) TRAVIS J.
 PA (POTE/) POTEMPA J.
 PA (BANB/) BANBULA A.
 XX Travis J, Potempa J, Banbula A;
 PI WPI; 2000-594181/56.
 DR Prolyl tripeptidyl-peptidase, active analog, fragment or variant useful
 PT

PT for identifying its inhibitor which is useful for protecting an animal
 PT from a periodontal disease such as gingivitis and periodontitis
 XX Claim 22; Fig 6; 58pp; English.

CC The present sequence represents a H2 homologue of a prolyl
 CC tripeptidyl-peptidase (designated DPP) from Porphyromonas gingivalis.
 CC The prolyl tripeptidyl-peptidase has an amidolytic activity, and cleaves
 CC a peptide bond in a target polypeptide having at least 4 amino acids.
 CC This bond is between a proline and an amino acid attached to the
 CC alpha-carboxyl group end of the proline. The polypeptide is useful
 CC for identifying inhibitors. These inhibitors are then useful for
 CC reducing the growth of bacterium or for protecting an animal from a
 CC periodontal disease such as gingivitis and periodontitis caused by
 CC Porphyromonas gingivalis.

SQ Sequence 841 AA;

Query Match 13.0%; Score 453; DB 21; Length 841;
 Best Local Similarity 23.5%; Pred. No. 2.2e-29;
 Matches 174; Conservative 121; Mismatches 291; Indels 154; Gaps 29;

QY 1 MKKLITADITAA--IVSVTDFOYAPD-----GT-----RA 28
 Db 157 lkslllegdtatdrvrvvkpkrtardssalypnytgkerlsikhmmsgtflsggsalsptg 216
 QY 29 AYKSOVNOEKDSYTSNIWIYETKTGSGVPWTHGKERTDPWSPDGRTLAFISREGDA 88
 Db 217 kylvitsyrsvrdnkpavtynqlrdakgnlllnkealg--wmpbedlmvirk--kegna 273
 QY 89 AOLYIMTEG--GEARKLTDIPYGVSKPLWSPDGESIL-VTISLGE-----SIDDR 138
 Db 274 krlvafpdmgkextlvsnlp--esqfmspdaryylfykqekgpgkplfirhldpdr 331
 QY 139 EKTEDSYEPVEVVOGLSYKRDGKGLTRGAYAOQLVSVKSGEMKELTSHKADHGDPASP 198
 Db 332 qsdwdr-----sqilynaesgvypgtlftgysttyldiap 368
 QY 199 DGRWLVSANLTETDDASKPH---DVIYMSLESGDLKQVTPHRGSGSSSPDGRYAL 255
 Db 369 dskraligt--lstdwtrrpfrrfatlmeymetgkadtlltrdpsidaigytpdkhli 426
 QY 256 LGNEKEYKNATLS-----KAWLYDIEOGRUCLTEMLDVHLADALIGSLIGG 303
 Db 427 mgsadaifgnlgnlksvtpnsydkqffldlstrkataltkfn-----psvsag 477
 QY 304 AEORPIWTKDSQGFYVIGTDCGS--TGIVYISIEGL-VYPIRLEKEYINSFSLSPDEOHFI 361
 Db 478 rfr-----knnnyyfraengsrkqlyrldlktleisqigtgedvqvfwgvaadngav 531
 QY 362 ASVTKPDRPSELYSIPLGQEEKOLTGANDKFV---REHTISI---PEIQYATEDGV 413
 Db 532 ysgqsannadrly-----rldgtkgklvwdlsaeklanldftpardwnytapdgtv 582
 QY 414 VNGWLMRPAQMEGETTYPILNTHGGP---HMVYGHYTFHEQVLAAGYAVVYINPRG 469
 Db 583 vegwylppqfpskkykypmlvyvggtsinrtieg---yslamyaaggyvvytlnpsg 639
 QY 470 SHYGQEF-----VNAVRYGDKYDDVMQAVDEAIKRDPHIDPKRLVGTGSGYGGFM 525
 Db 640 ttgyggyaahvna-----wgdrtadeiigatkefirthsfvngkvgcfasyggyfmg 695
 QY 526 WIVGOTNRFKAATVQORSINWISFHCVSIDIGYFTDQLEHDMFDETEKL-WD----- 577
 Db 696 yltktddifaavaahagissisnywg---sgy-----wgmgystvastdsypnnpdyag 748
 QY 578 RSPLKYAANVETPLLIHGERDRRCPIEQAEQLFKKMGKETKLVRFNASHNLSRTG 637
 Db 749 hspflradkhtplllhgsvntnvtasvnlalnalkilgrevefietgdhfil---- 805
 QY 638 HPRQRIKRLNYSISSWFOHL 657
 Db 638 HPRQRIKRLNYSISSWFOHL 657

Wed Sep 26 11:29:27 2001

us-09-462-845-2.rag

806 eperirwtnsicawfarwl 825

RESULT 7

AAP94265
ID AAP94265 standard; protein; 783 AA.

XX AC

XX AC AAP94265;

XX DT 24-JUN-1990 (first entry)

XX DE Sequence of APH36.1 clone.

XX KW Clone APH36.1; acyl-peptide hydrolase.

XX OS Rattus rattus.

XX FH Key Location/Qualifiers

FT Protein 1..721

FT /label=claimed protein 580..587

FT Active-site 130.

FT Region /note="potential site of glycosylation" 131.

FT Region /note="potential site of glycosylation" 132.

FT Region /note="potential site of glycosylation" 229.

FT Region /note="potential site of glycosylation" 230.

FT Region /note="potential site of glycosylation" 231.

FT Region /note="potential site of glycosylation" 239.

FT Region /note="potential site of glycosylation" 240.

FT Region /note="potential site of glycosylation" 241.

FT /note="potential site of glycosylation"

XX EP303997-A.

XX PD 22-FEB-1989.

XX PE 13-AUG-1988; 88EP-0113186.

XX PR 21-AUG-1987; 87US-0087936.

XX FA (GEHO-) THE GENERAL HOSPITAL CORP.

XX PT Smith JA;

XX WPI; 1989-055444/08.

XX N-PSDB; AAN91052.

XX Recombinant acyl-peptide hydrolase -used to catalyse hydrolysis of N-acyl peptide(s) or reaction of N-acetyl amino acid donor and acceptor protein.

XX Disclosure; : 2lpp; English.

XX Sequence contains the claimed sequence of acyl-peptide hydrolase (APH) (claim 1, page 11). APH can be used to catalyse the hydrolysis of the N-terminal acyl amino acid of an acylated polypeptide, or the reaction between a derivatised N alpha-acetyl amino acid donor and acceptor with a free alpha-NH2. It can also be used to make refractory proteins susceptible to Edman sequencing or th reduce degradation of proteins to be used therapeutically.

XX Sequence 783 AA;

XX Query Match 11.4%; Score 396.5; DB 10; Length 783;

XX Best Local Similarity 21.7%; Pred. No. 1.1e-24;

XX

Matches 148; Conservative 126; Mismatches 271; Indels 137; Gaps 21;

QY 22 ADGTRAAVVK---SOVNOEKDSYTSNIW-----IYETKGGSPVW 59
Db 93 spsgtmkavlirkaggtvgeekqfl-evweknrkklksfnlsalekhgpyveddcfcgclsw 151
QY 60 THGEKRSPTDPRWSPDGRTLAFISDRGDAALYIMTEGGEARKLTDIPGVSKPLWSPD 119
Db 152 shseth-----llyvaeek-----rpK 168
QY 120 GESILVTISLGESESIDREKTEQDSVEPVEVOGLSVKRDGKGLTRGAYAQQLVLSVKSG 179
Db 169 aesffqtkaIdIsasddemarpkpkdaikdgfvfyedwgetmvsksipvlcvldids 228
QY 180 EMKEL--TSHKADHGDPAFSPDGKWLVSFA-----MLTETDDASKPHDVYINSLSESDL 231
Db 229 nisvlegvpennvsgqafwapgdtgvfvghwhepfirgiryctnrtsalysyvdslsgkc 288
QY 232 KQVTPHRGSGSSFSFDPGRYLALLGNEKEYKNATLSKAMLYDIEOGRLTCLTLMVDVHL 291
Db 289 ellisdgslalcspriLspdqcrivylgypclaphhqcslclyd----wytkvtvv-vdi 343
QY 292 ADALIGDSLIG-GAEQRPI--WTKDS-----QGFYVIGTDOGSTGIYIYSIEG 336
Db 344 vprqlgesfsgiyCSllplgcwsadsqrvvfdsaqrsrqdlfavdtqts--itsltaag 401
QY 337 LVYPIRLEKEYINFSLSPEQHFIAVTKPDRPSEL---YSIPLGOEE-----K 383
Db 402 sagswkl-----ltidklmvaqfstpslppsikvgflpppgkeqsvswsleae 452
QY 384 QLTGANDKFVREHTISIPETQYATGDMVNGWLMRPAQMEGETTYPLIINIHGGPHMM 443
Db 453 pipgihwgrvrlhppdqenvqyadldfeai---llqpsnpdktpvmvmpghgphss 509
QY 444 YGHTYFHEFQVLAAGYAVVYINPRSHGYGOEFVNAVVRGDKYDDVMQAVDEAIKR 503
Db 510 fvtawmlfpamckmgfavllyvrgstgfgdsillsipgnvghqdkdvqfaveqvige 569
QY 504 DPHIDPKRLKLVGGSYGGFMTNWIWGO--TNRFKAATORSISNWSIFHGVSDIGYFFFDW 562
Db 570 e-hfdaarvalmgshghgflschli9qypetysacianrpviniasmmgstidi---pdw 624
QY 563 QLEHDMFEDT-----EKLMDRSLPKYAANVETPLLILHGERDRCPTEQAEQLFI 612
Db 625 cmvetgfyfysnsclpdlnvweemldksplkyipqvktpvllmqlgqedrivrpfkqgmeyyr 684
QY 613 ALKMKKETKLVRFPPNASHNLS 634
Db 685 alkarnvprvrllypkshnals 706

RESULT 8

AAR11056
ID AAR11056 standard; Protein; 732 AA.

XX AC

XX AC AAR11056;

XX DT 23-MAY-1991 (first entry)

XX DE Acylamino acid-isolating enzyme-like polypeptide.

XX KW AARE; pig liver.

XX OS Sus scrofa.

XX PN JP03030673-A.

XX PD 08-FEB-1991.

XX XX 29-JUN-1989; 89JP-0165216.

XX XX 29-JUN-1989; 89JP-0165216.

XX PR

PA	(TAKA-) TAKARA SHUZO KK.	
XX		
DR	WPI; 1991-084340/12.	
DR	N-PSDB; AAK10958.	
XX		
PT	Acylamino acid-isolating enzyme-like polypeptide - prepd. by	
PT	genetic engineering.	
XX		
XX		
PS	Claim 1; Fig 2; 10pp; Japanese.	
XX		
CC	The sequence was obtd. from six clones isolated from a pig liver	
CC	cDNA library. The clones, lambda AARE 419, 450, 451, 452, 521 and	
CC	522 cover the entire sequence of the gene.	
XX		
SQ	Sequence 732 AA;	
	Query Match	11.2%; Score 391; DB 12; Length 732;
	Best Local Similarity	21.0%; Pred No. 3e-24;
	Matches 149; Conservative 123; Mismatches 287; Indels 152; Gaps	23;
QY	22	ADPGTAAAVKSO--VNQEKDSYTSNIW-----IYFTKGSVPWT 60
DB	97	spsgmkavlrkaggtgtaeekqflevweknrkfrlsalekhgpyveddcfgcls 60
QY	61	HGEKRSTDPRMSPDGRITLAFTISDREGDAQAQLYIMSTEGEGARKLTDIPYGVSKPLWSPDG 120
DB	157	hseht-----llyvaek-----rpka 173
QY	121	ESILVTISLGEGESIDDRKTEQDSYEPVEVQGLSYKRGDKGLTRGAYOALVLVSVKSGE 180
DB	174	esfftkaldvtsgsdemartkkpqdqlkgdqflfyedwgenmvskstpvlcvidiesgn 233
QY	181	MKEL--TSHKADHGDPAPSPDGKWLVSFA-----NUTETDASKPHDHYIMSLSGDLK 232
DB	234	isvlegvpesvppqafwaptgdvtyvfgvwhepfrlgrfctarrsalyvvdltgkce 293
QY	233	QVTHRGSGFGSSFSPOGRYLALGLNEKEYKNATLSKAWLYDIEQGRLTGLTEMLDVHLA 292
DB	294	llsdesavtspripdqcrivylrfpslvphqcgqlcyd---wytvtsvv-vdiv 348
QY	293	DALIGDSLIG-GAQRPI--WTKDS-----QGPYVIGTDQGSTGLYIISTEGL 337
DB	349	prlgedfsglycslplpgcwsadsqrvfdsprsrqdlfavdtgmgsvtstaggsgg 408
QY	338	VPIRLKEYINSFSLSPDQHFHTASVTKPDRSEL---YSIPLGOEEKOLTANDKFKVR 394
DB	409	swkl-----ltidrlmvvqfstpsvpslkvgflpagkeq-----avswsl 452
QY	395	EHTISIP-----FEIQYATEDGVNVMGLMRPOMCEGTYTPIILNIHG 439
DB	453	eeaeppdliswslrvlpppqghevqya---gldfaillqpsnspekqlqpmvmpghg 509
QY	440	PHMYGHTYFHEFQVLAAGYAVVYINPRGSHGCGYGEFFNAVRGDYGKDYDDVMQAVDE 499
DB	510	phsftvawmlfpamckmgfavllyvnyrgstfgqdsilsipgnvghqdvkdvfaveq 569
QY	500	AIKRDPHIDPKRLGVYGGSGGPMWNIYVQ--TNRFKAAVTRQSRISNWSIFSHGVSDIGYF 558
DB	570	vigee-hfdagrvalmgshggfllschllgqypetysacvvpnpvlniasimgstdi--- 625
QY	559	FTDQLEHDMFEDTE-----KLW----DRSPLKYAANVETPLLILHGRDRDRCPTEQAE 608
DB	626	-pdwcmveagfysyddclpdlswaamlckspkyapqvktpllmlggedrrvpfkqgm 684
QY	609	QLFIALKMGKETKLVRFNPASNHLNLSRTGHPQRKRLNIYSS--WFOQHL 657
DB	685	eyyrvikarnvpvrilllypksthalsev-----evesdsfmnavlwlclthl 730

RESULT 9

AAK20001

ID AAR20001 standard; Protein; 732 AA.

[illegible]

RESULT 9
AAR20001
ID AAR20

XX PA (DAND) DANA FARBER CANCER INST INC.
 XX PI Morimoto C, Schlossman S, Tanaka T;
 XX XX WPI; 1994-151317/18.
 XX PT Polypeptide fragments and analogues of CD26 and encoding nucleic
 XX PT acid - useful for stimulating immune response, e.g. for treatment
 XX PT of AIDS to counteract immunosuppressive drug, and as vaccine
 XX PT adjuvant.
 XX PS Claim 3; Page 49-52; 85pp; English.
 XX XX
 CC The sequences given in AAR54612-14 represents analogues of the human T
 CC cell activation antigen CD26 which have internal deletions. The
 CC analogues pref. lack residues 3-9 or 24-34. These analogues are
 CC soluble under physiological conditions and lack enough amino acid
 CC residues to render them susceptible to cleavage by signal peptidase.
 CC The peptide fragments and analogues are useful as immune or response-
 CC stimulating therapeutics, eg. they may be used for treatment of
 CC disease conditions characterised by immunosuppression, eg. AIDS or
 CC AIDS-related complex, other virally or environmentally-induced
 CC conditions, and certain congenital immune deficiencies. The peptides
 CC can be employed to increase immune function which has been impaired by
 CC use of immunosuppressive drugs, such as certain chemotherapeutic drugs.
 XX XX
 SQ Sequence 759 AA;

Query Match 7.8% Score 272; DB 15; Length 759;
 Best Local Similarity 20.4%; pred. No. 3.6e-14;
 Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;

QY 15 SVTPQVAPDGTAAAYKVSQVQKDSYTSNIWY-----ETKGGVPTWTHGKR 65
 Db 94 sindyspsdqfilleynvkvhrsyasydyidinkqliteeripnntqw----- 148
 QY 66 STDPWSPDGRITAFISDRGDAQLYIMSTEGEARKLT-----DIPY----- 109
 Db 149 ----twspvghklaywnnd-----iyvkieplnpsryitwtgkediiyngitdwvyeee 199
 QY 110 ----GVSKPLWSPGESILVTISLGESESIDRE--KTEQDSYEPVEVQGLSKYRDKGLT 164
 Db 200 vfsaysalwspngtflaya-----qfndtevpilleyfsydesiq----- 240
 QY 165 RGAYAQLVLVSV-KSGEMKELTSHKADHGPAPSPGKWLWVFSANLTETDDASKPHDVI 223
 Db 241 ----yptktrvypypkag-----avnptvkffv-----vntdsls----- 270
 QY 224 MSLESGDLKQVTPHRGSGSSPSPDGRYLALLGNEKEYKNAT---LSKAWL----- 272
 Db 271 -svtnatsiqit-----apasmliqdhylc-----dvtwatqerisliqlriqnysv 317
 QY 273 ----YDIEQGRLTCLTEMLDVHLADALIGDSLIGG-AEQRPITWKDSQCFY-VIGTDOG 325
 Db 318 mdicdydessgrwncvlvarqhiemstt-----gwwgrfrpsephftldgnsfykiisneeg 373
 QY 326 STGIYVYISIE--GLVYPIRLEKEYINSFSLSPDEQHFIAVTK----- 366
 Db 374 yhicfyqiddkdcfitgtwviegiealtsdylyisneykmpggrnlykqisldyt 433
 QY 367 -----PRPSELVSIPLGOPEK--QL-----TGANDKFVR--EHTIS 399
 Db 434 kvtcslcelnper-cqyysvsvsfskaeyqlrcsgpglplytlhsrvndkgirvlednsa 492
 QY 400 IPTEEIYQATE-----DGVWVNG---W--LMRPAQMEGETTYPLIINHGPHMYGHTYF 449
 Db 493 idklmqnvqmpskkldifilnetktywgmilpphfdkakkypdllldvyagpcsqkadtfv 552
 QY 450 H---EFQVLAAGYAVVYINPRGSHGYGQEFVNAVVGVDYGGKDXDDVQWQVDEATKRDPH 506
 Db 553 rlnwatylastenlivasfgrgsgyqgdkimhainrrlgtfevedqieaarqfsmkg-f 611

QY 507 IDPKRLGVGTGSGYGFMTNIIWGO-TNRFKAAYTORISINWISFHVSDIGVFFTDWOLE 565
 Db 612 vdnkriaiwsgsygyvvtsmvlgsgsvfkcgiavapvsrweyysv-----yterymg 665
 QY 566 HDMFEDTEKULWDRSPLKYAAN--VETPLLILHGERDDRCPIEOAEQLFIALKKMGKETKL 623
 Db 666 lptpednlldhrynstmsraenfakveyllingtadnvnifqgsaqiskalvdvgvdfga 725
 QY 624 VRFPNASHNL-SRTGHPQRQRIKRLNYISSWF 653
 Db 726 mwytdedhgiastahqhiythmshfikqcf 756

RESULT 12
 AAR40909
 ID AAR40909 standard; Protein; 766 AA.
 XX
 AC AAR40909;
 XX
 DT 05-FEB-1994 (first entry)
 XX
 DE Sequence encoded by human CD26 cDNA.
 XX
 KW Human T cell activation antigen; monoclonal antibody Tal.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH Region 7..28
 FH FT /label= hydrophobic
 FH Region 29..323
 FH FT /label= N-terminal glycosylated region of
 FH FT extracellular domain
 FH FT /note= "8 sites for N-linked glycans"
 FH FT 324..551
 FH FT /label= Cysteine rich region of extracellular
 FH FT domain
 FH FT /note= "1 N-linked glycosylation site"
 FH FT 552..766
 FH FT /label= C-terminal region of extracellular
 FH FT domain
 FH FT /note= "1 N-linked glycosylation site & 1
 FH FT catalytic site"
 FH FT 627..631
 FH FT /label= active site of serine protease/esterase
 FH FT /note= "fits the consensus sequence GXSGX"
 XX
 PN W09316102-A.
 XX
 PD 19-AUG-1993.
 XX
 PF 09-APR-1992; 92WO-0502892.
 XX
 PF 06-FEB-1992; 92US-0832211.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 XX
 PI Morimoto C, Schlossman SF, Tanaka T;
 XX
 DR WPI; 1993-272827/34.
 DR N-PSDB; AAO46089.
 XX
 PT Polypeptide fragments of CD26 - are capable of disrupting binding
 PT of CD45 and CD26 and thus interfering with T-cell activation
 XX
 PS Disclosure; pages 39-43; 73pp; English.
 XX
 CC C26 is a human T cell activation antigen originally identified by
 CC its reactivity with the MAb Tal. C26 cDNA library was constructed
 CC from human PHA-activated T cells using the CDM/vector. The hydrophobic
 CC N-terminal of the predicted CD26 polypeptide has the characteristics
 CC of a signal sequence of the type II membrane protein, which is

Db 325 mdicdydessgrwnciavarqhiemstt---gwgfrpsephftldgnsfykiisneeg 380
 QY 326 STGIYVISTE--GLVYPIRLEKEYINSFSLSPDEQHFIASTVTK----- 366
 Db 381 yrhicfyqiddkdwftikgtewigiealtsdylyisneykgmpgggnlykiglsdyt 440
 QY 367 -----PDRSELVSIPLQOEK--QL-----TCANDKFRV--EHTIS 399
 Db 441 kvtslscelnper-cqyyvsfskskaekyyqlrcsgpgpilytlhsvndkgirvledhsa 499
 QY 400 IPEEQYATE-----DGVWVNG---W---LMRPAOMEGETTYPILININGGPHMYGHTYF 449
 Db 500 idkmlnqvmpskklldfillnetkwygmilpphpfkdkkypplldvayagpcsqkadtvf 559
 QY 450 H---EFQVLAAGYAVVYINPRGSHGYGOEFVNAVVRGVDYGGKDDVDMQAVDEAIKRDPH 506
 Db 560 rlnwatylastenilivasfdrsgyqgdkimhainrrlgfvevedqleaargfsmg-f 618
 QY 507 IDPKRLGVGTGSGYGFNTWIVGO--TNRFKAAVTQRTSISNWSIFHGVSDIGYFFTDWOLE 565
 Db 619 vdnkrlalwgsygyvtsmvlsgsgvfkcgiaavpvrweyydsv-----yterymg 672
 QY 566 HDMPEDETEKLDWRGPLKYAAN--VETPLLLHGERDDRCPIEQAEQFLIAKKMGKETKL 623
 Db 673 lptpednlhynrstvmsraenfkyveylilngtaddnvnfhqgsaqiskalvdvgvdfqa 732
 QY 624 VRPNASHNL--SRTHGPRQRIRKRLNYSISWF 653
 Db 733 mwytdedghasstahqhiythmshfikqcf 763

RESULT 14

AAW89596 standard; Protein; 931 AA.

XX AAW89596;

XX 17-MAR-1999 (first entry)

XX Saccharomyces cerevisiae dipeptidyl aminopeptidase I.

XX Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;

XX KW flavour enhancer; palatability; mouthfeel; aroma; crust colour;

XX KW baking; animal feed additive; hydrolysis.

XX OS Saccharomyces cerevisiae.

XX PN WO9851803-A1.

XX PD 19-NOV-1998.

XX PF 12-MAY-1998; 98WO-US09629.

XX PR 20-OCT-1997; 97US-0062892.

XX PR 16-MAY-1997; 97US-0857884.

XX PA (NOVO) NOVO NORDISK BIOTECH INC.

XX PI Blinskovsky A, Brown K, Byun T, Klotz A, Rey MW;

XX DR WPT; 1999-045232/04.

XX PT New dipeptidyl aminopeptidase from Aspergillus oryzae used to

XX PT produce protein hydrolysates enriched in particular amino acids,

XX PT useful as flavour enhancers, e.g. in doughs

XX PS Example 7; Page 59-61; 77pp; English.

XX CC The present sequence represents dipeptidyl aminopeptidase (DPAP) from

XX CC Saccharomyces cerevisiae. DPAP acts synergistically with aminopeptidase

XX CC (AP) to hydrolyse polypeptides, producing protein hydrolysate (PH),

XX CC useful in foods as flavour enhancer, e.g. in baked goods, enriched in:

CC (a) Ala, Arg, Asp, Gly and/or Val, or (b), if the substrate has been
 CC deamidated, in Glu (free and/or peptide bound), in which case products
 CC are useful as animal feed additives. DPAP can also be used in flavour-
 CC improving compositions (optionally containing AP) and in dough pre-mixes,
 CC also for deactivating enzymes and for converting precursors to mature
 CC proteins. DPAP increases the level of hydrolysis of proteins and thus
 CC of flavour development, and a mixture with AP may hydrolyse tripeptides
 CC that are resistant to either enzyme used alone. PH have improved
 CC solubility, emulsifying and foaming properties, and products containing
 CC them have better flavour, palatability, mouthfeel, aroma and crust
 CC colour.

XX Sequence 931 AA;

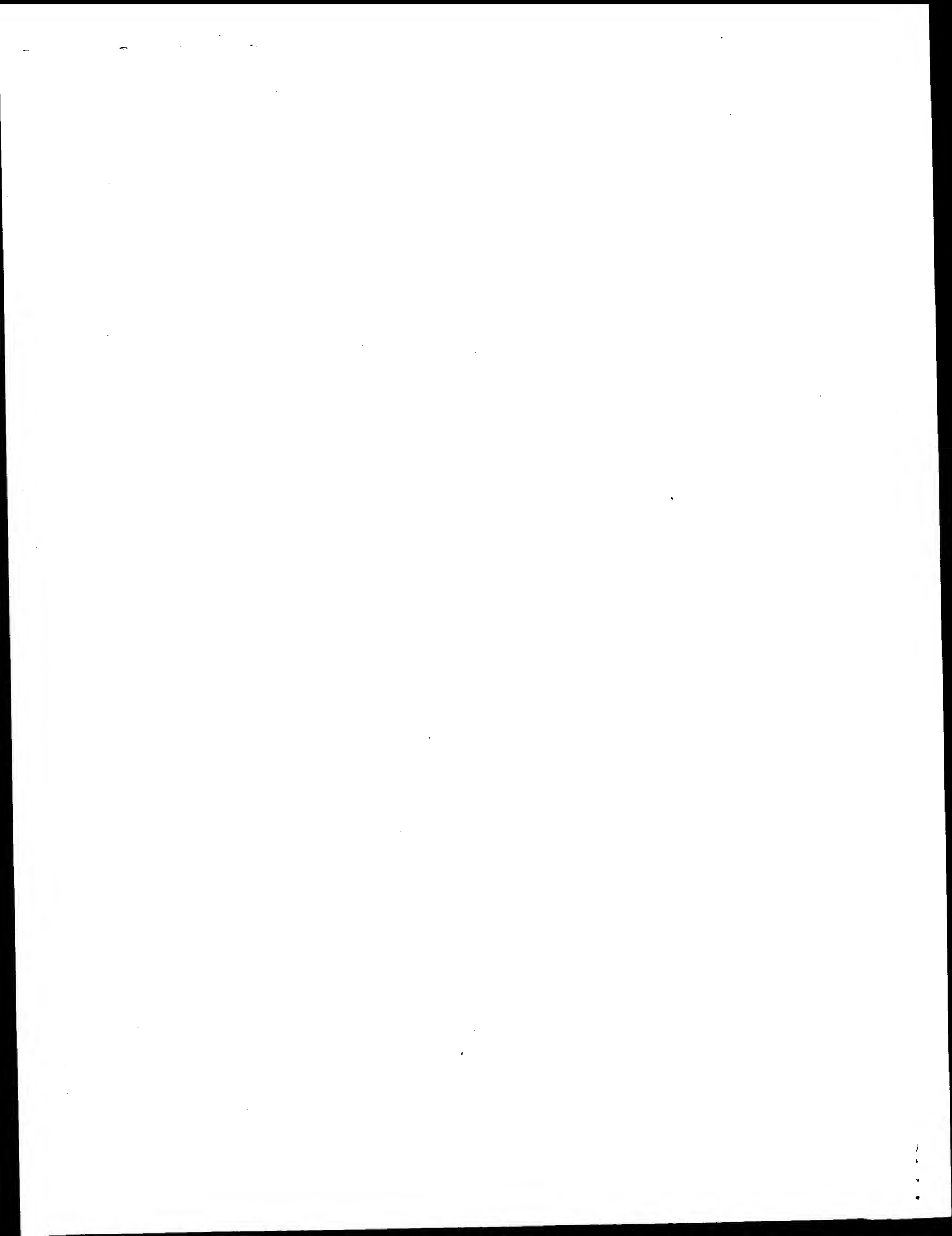
Query Match

Best Local Similarity 20.1%; Score 269; DB 20; Length 931;

Matches 155; Conservative 108; Mismatches 298; Indels 212; Gaps 30;

QY 9 DITALVSTDPQYAPDGTTRAAAYVKSOVQOEKDS--YTSNI-----WIYETKNG 54
 Db 220 detfevnlggnrflyegveftvstqinykldklifgtlesefthsskgfywikdlnctg 279
 QY 55 GSVPTWTHGEKSTD-----PRWSPDGRTLAFISDREGDAALQYIMSTEGGEARKL 104
 Db 280 niepilpekaddhneyelglsklsyahfspaynyfyfven-----nlflqgvnsqvakv 334
 QY 105 TDIPYGVSKPL-----WSPDGESILVTISLGGESIDDDREKTEQ 143
 Db 335 te---dgskdifnakpdiyevevlasdaqaiwapddsk--avfarfnadtvsddirlary 389
 QY 144 DSYEPVEVQGLSYKRDGKGLTRGAYAOVLVSVKSGEMKEL--TSHKADHGDPAFSPDGKW 202
 Db 390 tmneaylsdtkikypkpgfqnpqf-dlflvnlngliysintgqkd----silyngkw 444
 QY 203 L---VFSANLTETDDASKPHDVYIMSLGSG-----DLKQVTP-- 236
 Db 445 isptdrfrefi--tdrnskildkvvydipssqmltrvntnsnlfngwtektkdiisipkp 502
 QY 237 --HRGFGSSSPDPGRYLALL-----GNEKEYKNATLSKAWLYDIEQG 278
 Db 503 elkrmdygyidihadsrgfshfyptvfakepiqltkgnwvgtngivg----yeyetd 558
 QY 279 RLCLTEMLDV---HLADALIGDSLIGGAFQRPILWTKDSOGFVYIGTDOGSTGIYYSIE 335
 Db 559 tiffaneigvmsqhlhysisltsttqntfqsignpsskdyfy----- 601
 QY 336 GLVYPIRLEKEYINSFSLSPDEQHFIASTVTKPDRP-----SELYSIPLGQEE 382
 Db 602 -----dfelssaryaiskklgtpdkvqpltrvlnvaihdsilqlt 647
 QY 383 KOLTGANDKF---VREHTTISPEEIQYATEDGVWVNGWLMRPAOMEGETTYPILINHGG 439
 Db 648 k-----dekfkekiknydipitsyktmviddveinyieikpanlnpkkypilvnygg 702
 QY 440 PHMYGHTYF-----HEFQVLAAGYAVVYINPRGSHGYGOEFVNAVVRGDDGGKDYDD 492
 Db 703 p-----gsqfttksslafeqavvsgldvliqieptgctgkgwsfswareklygweprd 758
 QY 493 VMOQAVDEAIKRD--PHIDPKRLGVGTGSGYGFMTNIVGQTN--REKAAVTQRTSISNWSIF 549
 Db 759 itevtktkfignsqhideskiaiwgwsyggfslktveldngdtfkyamavapvntwtly 818
 QY 550 HGV-----SDIGYFFTDWOLEHDMFEDTEKLMDRSPKLKYAANVETPLLHGR 598
 Db 819 dsvyterymnapsenhegyfsvstignfksfslrlf-----ivhgtf 862
 QY 599 DDRCDIEQAEQIETALKKMG--KETKLVRFPNASHNLSRTGHPQRORI--KRLNRY 648
 Db 863 ddnvhigtfrlvdqnlilngltnymhifpdsdhsiry--hnaqirivfqklyy 913

RESULT 15



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 17:35:13 : Search time 25.92 Seconds
(without alignments)
1930.813 Million cell updates/sec

Title: US-09-462-845-2

Perfect score: 3489

Sequence: 1 MKKLITADDITAIIVSVTDPO.....HPQRKIKRLNISSWFQDHL 657

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	3489	100.0	657	2 E70025	probable acylamino
2	1104.5	31.7	642	2 C71137	hypothetical prote
3	1096.5	31.4	631	2 F75007	probable acylamino
4	995	28.5	655	2 E75551	probable acyl-pept
5	962.5	27.6	659	2 F72568	probable acylamino
6	903	25.9	674	2 B84381	acylaminoacyl-pept
7	697	20.0	572	2 F72455	probable acylamino
8	667.5	19.1	632	2 E75057	peptidase PAB1418
9	660	18.9	622	2 F71174	hypothetical prote
10	473	13.6	709	2 B82580	alanyl dipeptidyl
11	410	11.8	591	2 H72474	probable acylamino
12	405.5	11.6	721	2 T09631	probable acylamino
13	396.5	11.4	732	1 S07624	acylaminoacyl-pept
14	391	11.2	732	1 J00132	acylaminoacyl-pept
15	384	11.0	732	1 UC4655	acylaminoacyl-pept
16	357.5	10.2	741	2 JC5142	X-Pro dipeptidyl-p
17	355.5	10.2	598	2 F84199	hypothetical prote
18	350.5	10.0	745	2 T33751	hypothetical prote
19	346	9.9	637	2 S75772	probable peptide h
20	334.5	9.6	606	2 T35378	dipeptidyl-peptida
21	328.5	9.4	795	2 F82858	probable acylamino
22	319.5	9.2	569	2 S74053	F44B9.1 protein -
23	319.5	9.2	761	2 T15907	hypothetical prote
24	303	8.7	629	2 T15945	probable acylamino
25	287.5	8.2	582	2 D72636	X-pro dipeptidyl-p
26	278	8.0	711	2 S66261	dipeptidyl-peptida
27	272	7.8	766	1 CDH026	dipeptidyl aminope
28	269	7.7	931	2 A49737	dipeptidyl-peptida
29	243	7.0	760	1 S23752	dipeptidyl-peptida

ALIGNMENTS

RESULT 1

E70025

probable acylaminoacyl-peptidase (EC 3.4.19.1) yuxL - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: E70025; D25364

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: E70025

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-657 <KUN>

A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15213.1; PID:el1843

A:Experimental source: strain 168

R:Parrot, C.

EMBO J. 5, 3013-3019, 1986

A:Title: Evolution of biosynthetic pathways: a common ancestor for threonine synthase

A:Reference number: A91055; MUID:87080286

A:Accession: D25364

A:Molecule type: DNA

A:Residues: 1-MPSVPORRAPFFAGDKGVLTNTCKGDEVE', 3-17, 'S', 19-24 <PAR>

A:Cross-references: GB:X04603; NID:g40210; PIDN:CAA28272.1; PID:g40213

A:Note: hypothetical fragment that is probably a mistranslation of a portion of the y

C:Genetics:

A:Gene: yuxL

C:Keywords: hydrolase; omega peptidase

F:516,599,631/Active site: Ser, Asp, His #status predicted

Query Match 100.0%; Score 3489; DB 2; Length 657;
Best Local Similarity 100.0%; Pred. No. 8.le-226;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKKLITADDITAIIVSVTDPOYAPDGTAAAYVKSQVNOEKDSYTSNIIYETKGSVPWT 60
|||||
Db 1 MKKLITADDITAIIVSVTDPOYAPDGTAAAYVKSQVNOEKDSYTSNIIYETKGSVPWT 60
|||||
Qy 61 HGEKRSTDPKWSPDGRTLAFISDREGDAALYIMSTEGGEARKLTIDIPYGVSKPLWSPDG 120

61 HGERSTDPWSPDGR^TLAFISDREGDAAQLYIMSTEGGEARKLTIDIPYGVSKPLASPDG 120
|||||
121 ESILVTISLGGESIDDRKTEQDSYPPEVVOGLSYRKDGKGLTRGAYAOALVLVSVKSGE 180
|||||
122 ESILVTISLGGESIDDRKTEQDSYPPEVVOGLSYRKDGKGLTRGAYAOALVLVSVKSGE 180
|||||
181 MKELTSKHADHCDPAFSDPGKWLVFSANLTTETDASKPHDVYIMSLSEGDLLKVTPHRGS 240
|||||
182 MKELTSKHADHCDPAFSDPGKWLVFSANLTTETDASKPHDVYIMSLSEGDLLKVTPHRGS 240
|||||
241 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMILDVHLADALIGDSL 300
|||||
242 FGSSSFSPDGRYLALLGNEKEYKNATLSKAWLYDIEQGRLTCLTEMILDVHLADALIGDSL 300
|||||
301 IGAQEORPIWTNDSQGFYVIGFDQSGTGIIYSIBGLVYPIRLEKEYINFSLSLSPDEQHF 360
|||||
302 IGAQEORPIWTNDSQGFYVIGFDQSGTGIIYSIBGLVYPIRLEKEYINFSLSLSPDEQHF 360
|||||
361 IASVTKPDRPSELYIPLGOEEKQLTGANDKFVRHTTISIPEIQYATEDGMVNWGLMR 420
|||||
362 IASVTKPDRPSELYIPLGOEEKQLTGANDKFVRHTTISIPEIQYATEDGMVNWGLMR 420
|||||
421 PAQMEGETTYPLTLNTIHGGPHMMYGHTYPHEFQVLAAGYAVVYNPRGSHGYGOEFVNA 480
|||||
422 PAQMEGETTYPLTLNTIHGGPHMMYGHTYPHEFQVLAAGYAVVYNPRGSHGYGOEFVNA 480
|||||
481 VRGDYGGKYDDVMQAVIDEAIRKDHPHDPKRLGVTGGSYGGFMTNWVGOTNRFAAAVTQ 540
|||||
482 VRGDYGGKYDDVMQAVIDEAIRKDHPHDPKRLGVTGGSYGGFMTNWVGOTNRFAAAVTQ 540
|||||
541 RSISNNWSIFHGVSIDIGYFFTDWOLEHDMFEDTEKLWDNRSPDKYAANVEPTLLILHGERDD 600
|||||
542 RSISNNWSIFHGVSIDIGYFFTDWOLEHDMFEDTEKLWDNRSPDKYAANVEPTLLILHGERDD 600
|||||
601 RCPTEQAEOLFIALKKMGKETKLVRRFPNASHNLRTGHPRQRIRKRLNYISSWFDOHL 657
|||||
602 RCPTEQAEOLFIALKKMGKETKLVRRFPNASHNLRTGHPRQRIRKRLNYISSWFDOHL 657
|||||

RESULT 2
-71137
Proteobacteria; Gammaproteobacteria; Alphaproteobacteria; Rhodospirillum rubrum; Pyrocytophage-like protein PH0863 - Pyrococcus horikoshii
Species: Pyrococcus horikoshii
Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
Accession: C71137
Author: Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushidomae, S.; Nakamura, K.; Nakamura, A.
Title: Complete sequence and gene organization of the genome of a hyperthermophilic bacterium, Pyrococcus horikoshii strain OT3
Reference number: A71000; PMID:98344137
Accession: C71137
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-642 <RAW>
Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29957.1; PID:g325727
Experimental source: strain OT3
Note: This accession replaces an interim accession for a sequence replacement.
C:Genetics
Gene: PH0863

145 SYPEVEVGLSVKRDCKGLTRGAYAQLVNLVSKSGEMKELTSHKADHGDPAFSPDGKWLIV 204
145 DVH--IIREIPFWNGVGIYKRNWVYLVDDVESGKKRLTPKNLNVDDQIRFH-NGR-LY 200
205 FSANITETDDASKP--HDVYIMSLGESDLKOVTPHRGSGSGSSP--DGRVLLALGNK 260
201 FTA--QEDREKPLISDLY--LENRKVRKLTTP--GKWRILDFLDDGSFV-LKANTL 252
261 EYKNATLSKAWLYDIHQGRLTCLTEMLDVLADALIGSLGGAEORPIWTKDSQGFYVI 320
253 ERGIPITNAHIYHDPKTKGELKLTCLDRNAYNSLNSD--VRGSOEALVYKGGWYIYV- 309
321 GTDQSGTGIYYSIEGLVVPRLRKEYINSFSLSPDEQHFIA--SVTKPDRPSELXIPL 378
310 ATDGPANLFRVNLDDKIERVITGGDRSVESFDIG-DYIAFTAQDAVT---PTELYIYRD 364
379 GQEEKOLTGANDKFVREHTISIPETQYATEDGMVNGWLMRPAQMEGETTYPLILNIHG 438
365 G-KKKVTDEN-KWIKGYTILSKSEHFKVKASDGEIDAWYMKPVNFRKGGKYPAILEIHG 422
439 GPHMYGHTYFHEFOVLAAAGYAAVYINPRGSHGYGOEFYNAVRGDYGKDYDDVMOAVD 498
423 GPKTAYGAYAEHHEHVLITSGFVIFSNPRGSDGYGEFAD-IRGHYGERDYQDLMEVVD 481
499 EAIKROPHIDPKLVGTGGYGGFMTNWIQVOTNRFRKAANTORSISNWIWSPHGYSDTICYF 558
482 EALRRFDFIDGERLGVGTGGYGGFMTNWIQVOTNRFRKAANTORSISNWIWSEFGTIDIGY 541
559 FTDWLEHDMFDEPEKLDWRSPLKAAANVETPLLIIHGERDDRCPIEQAEOLFALKMG 618
542 FAPDQIGKDPWSNLEGYWEKSPKAYPNVETPLLIIHSTEDYRCWLPEALQFLISLYLG 601
619 KETKLVFPNASHNLSTGTGHPRKRLKNLYLSSWFDOHL 657
602 KRVELAIFPGENHDLRSRGKPKRVRKRLIAGWKEWL 640
RESULT 3
7575007
Probable acylaminoacyl-peptidase (EC 3.4.19.1) PAB1300 [similarity] - Pyrococcus abyssi
A:Alternate names: acyl-peptide hydrolase (APH); acylamino-acid-releasing enzyme
S:Species: Pyrococcus abyssi
D:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 31-Mar-2000
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Molecule type: DNA
A:Residues: 1-631 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50502.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1300
A:Keywords: hydrolase; omega peptidase
A:Status: preliminary
A:Accession: H75007
A:Reference number: A75001
A:Description: Pyrococcus

Db 64 -P-DRCAWSPDGEIIFASRTVNSQNRKHVIAIVSPKGPQVQLHESVSPATS 116
 QY 114 PLWSPGESILVITISIGGESIDREKTEODSYEPVEVQGLSKYKRGKGLTRGAYAQVLV 173
 Db 117 IKWSSQSGMIGYLSRPTG-----REWKPYSERDVLDEIRIPVWFDESGWVFRYWGCLTV 171
 QY 174 VSVKSGE-MKELTSHKADHGDPAFSPDGKWLVSANLTETDDASKP--HDVYIMSLSESG 230
 Db 172 ISVPGGVLEKKGWVNVIVDFEAPDNTIVAV-----STDMMKPTIHLVLWDLASG- 226
 QY 231 LKQVTPHRG-SFGSSSPDGRLYALIGNEKEYKNATLSKAWLYDIEQRLTCLTEMLDV 289
 Db 227 -REKTLIEGLTIAAVFDPGRYIAVANDRKGLFSHYKIFVYDLNSEEFEVCLTCLDL 285
 QY 290 HLADALIGSLGAEQRPITWTKDSQGFYVIGTDOGSTGIYIIS-----IBGLVYPIRL 344
 Db 286 NTLATVNSDARGPSCLRWYW--DONGHLYGVHNGRVMVWYKSRPLGEAEAVLDP---S 340
 QY 345 KEYINSFSLSPDEQHFIASTVKPDRPSELYSIPLGQEEKQLTGANDKFKVREHTISIP 404
 Db 341 SATVDDFSISRGDTIAVVKMGTPSPDIY-ITYRDDNEYRLTDHNAWFAESRSRAEPVRL 399
 QY 405 QYATEDGVWNGWLMRPAQMEGETT--YPLILNTHGGPHMYGHTYPHEFOVLAAGYAV 462
 Db 400 QVQSPGLGSDAWILLPPD-AGECSGLPFWILYIHGGPKTSYGAFIHEFQMLASQGFV 458
 QY 463 VYINPRGSHGYGOEVNNAVRYGDKYDDVMOAVDEAIKRDPHIDPKRLGVTGSGYGGF 522
 Db 459 IYSNPRGSDGYSEFAD-LRGYGVDDYSELMKVVDLSDFDPQDQPGVGTGSGYGGY 517
 QY 523 MTNWIQVQTRNKAATVORSINWISFHVGSIDIGYFFTDWQLE-HDMFEDTEKLDWRSPL 581
 Db 518 MTNVIQVQTRNKAATVORSINWISFHVGSIDIGYFFTDWQLE-HDMFEDTEKLDWRSPL 577
 QY 582 KYAANVETPLILHGERDRCPTEOAEOLFALKKMGKETKLVRFNPNASHNLSRTGHPK 641
 Db 578 FSVENIETPLILHSTEDFCPLSGAIQLFTALKKGVETRLVFPNGENHLSRSGTPKR 637
 QY 642 RIKRLNYSWFDQHL 657
 Db 638 RVARLRAIASWFEHL 653

RESULT 6
 B84381
 acylaminoacyl-peptidase [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: B84381
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483
 A:Accession: B84381
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-674 <STO>
 A:Cross-references: GB:AE004437; NID:gl0581715; PIDN:AAG20414.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: yuxL

Query Match 25.9%; Score 903; DB 2; Length 674;
 Best Local Similarity 31.7%; Pred. No. 1.2e-52;
 Matches 223; Conservative 121; Mismatches 251; Indels 108; Gaps 22;

QY 5 ITADDTAIVSDPQYAPDGTAAVYKSVQVQEKDS--YTSNWIYETKTGSGVPTWTHG 62
 Db 4 VEAADYELARLSQAPVSGTDRVAVYVTRTAEEDGTQETASVHVAVPGVGGGSRRTAA 63

QY 63 EKSTDPNRWSPDGRITLAFISDREGDAA-----QLYIMSTEGGEARKLTDIPYGVSKPLWSP 118
 Db 64 DGTDAEPANWSPDGRITLAFISDREGDAA-----QLYIMSTEGGEARKLTDIPYGVSKPLWSP 122
 QY 119 DGSILVTISLIGGESID--DRE---KTEODSYEPVEVQGLSKYKRGKGLTRGAYAQVLV 168
 Db 123 DGRVAFLOQVPAERAEELDEHDEYERDTPDRVIDRQMYRGHTSYRD-----GAR 176
 QY 169 AOLVLVSVKSGEMKELTSHKADHGDPAFSPDG-----KWLVSANLTET 212
 Db 177 RHVYVADLAAGVRDRVTAGVAECGPGAWDGTLYYPIRRGLDADDRLEWAI----- 228
 QY 213 DDASKPHDVYIMSLGSDGLKQVTPHRSGSSSPDGRLYALIGNEKEYKNATLSKAWLYDIEQRLTCLTEMLDV 265
 Db 229 -EAATPAD-----SADSTVVTVEGM-----GPTLAVHGRSVAYTTTPADDP 271
 QY 266 TLSKAWLYDIEQRLTCLTEMLDVHLADALIGSLGAEQRPITWTKDSQGFYVIGTDOGSTGIYIIS 325
 Db 272 TQTAVEFEDRTGATDRLTGVRTVAERTAGHA-----PEWGPDEHLYVCTPDEG 323
 QY 326 -----STGIYVISTIEGLVPIRLEKEYINSFSLSPDEQHFIASTVKPDRPSELY-S 375
 Db 324 GYLVRVAADTGDDRVVLG-----DRHVAFSVSADAVAVVQS--EWDHPGEVNA 373
 QY 376 IPLGQEEKQLTGANDKFKVREHTISIPETQYATGDM--VNGWLMRPAQMEGETTYPLI 433
 Db 374 TPMGEEVTLTRANRDYLDERAVSKPESYMI---DGPAGDIQGVLTTPPEFDEETYP 430
 QY 434 LNHGCPHMY--GHTYPHEFOVLAAGYAVVYINPRGSHGYGOEVNNAVRYGDKYDDVMOAVDEAIKRDPHIDPKRLGVTGSGYGGF 491
 Db 431 VNVHGPPLHWSAAGTMMWHEFOTLAARGVVEVWCPNPGSTGYGOEVNNAVRYGDKYDDVMOAVDEAIKRDPHIDPKRLGVTGSGYGGF 490
 QY 492 DVMQAVDEAIKRDPHIDPKRLGVTGSGYGGFMTNWIQVQTRNKAATVORSINWISFHVGSIDIGYFFTDWQLE-HDMFEDTEKLDWRSPL 551
 Db 491 DVMQAVAAVTDTRD-YVDADNAFITTGSGFMTNWIQVQTRNKAATVORSINWISFHVGSIDIGYFFTDWQLE-HDMFEDTEKLDWRSPL 549
 QY 552 VSDIGYFFTDWQLEHDMFEDTEKLDWRSPLKYAANVETPLILHGERDRCPTEOAEOLFALKKMGKETKLVRFNPNASHNLSRTGHPK 611
 Db 550 TTD-AYALVEGEDTDPVRDNLAFWQAQSPAATDAVDTPTLVHSEADYRTPANTAEFLY 608
 QY 612 IALKMGKETKLVRFNPNASHNLSRTGHPRQIRKLNWISSWED 654
 Db 609 RLLKHGVDRVLYRFGEGHDLRSRGGDGHVVDRIERIAAWFD 651

RESULT 7
 F72455
 probable acylamino-acid-releasing enzyme APE2290 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: F72455
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Tanaka, Y.; Kudo, Y.; Tanaka, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
 A:Reference number: A72450; MUID:99310339
 A:Accession: F72455
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-572 <KAW>
 A:Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81302.1; PID:dl045088; PID:g
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE2290

Query Match 20.0%; Score 697; DB 2; Length 572;
 Best Local Similarity 33.5%; Pred. No. 5.9e-39;
 Matches 199; Conservative 99; Mismatches 248; Indels 48; Gaps 20;

QY 69 PRWSPDGRITLAFISDREGDAA-----QLYIMSTEGGEARKLTDIPYGVSKPLWSPDGE 124
 Db 69 PRWSPDGRITLAFISDREGDAA-----QLYIMSTEGGEARKLTDIPYGVSKPLWSPDGE 124

127 ISLGESEIDDRKTEODSEYEPVEVQGLSKYKRDGKGLTRGAYAQVLVSVKSGEMKELTS 186
Db 120 -----AVGFKREDEDFEDDVPWF---DNMGFFDGEKTFWFVIDTEGEVIE--- 166
QY 187 HKADHGDPAFSPDGK---LVFSANITETDDASKPH-----DVIYMSLESGLD 231
Db 167 ---QFEKPREP-SGIWHDGDSIVSV-----PHRDVPIFYFKYWDIYLW---KDGE 210
QY 232 KQVTPHRGSGSSSF---SPDGRYVALLGN-EKEYKNATLSKAWLYDIEOGRITCTLEML 287
Db 211 EKL-----FEKVSFYAIDSODGERILLYGPEKYYVSEH-DKIYIYD---GEVKGILDDI 260
QY 288 DVHLADLIGSLGGAGAEQPIWTKDSQGFYVIGTDQSGTGYIYISIEGLVYPIRLKEY 347
Db 261 DREVAQAKIRNGV-----YFTLFEEGSMNLYL---WDGEVREIAKGNHW 302
QY 348 INSFSLSPEDEQHFSTAVTKDPSPSELSIPLGQOEKQLTGANDKRVREHTISIPETQYA 407
Db 303 IMGEDA---DERLIYKLETATRAELYL---WDGEERQLTIDYNGLIFPKLKTFFEPHPRFK 357
QY 408 TEDGVYNGWLMRPAOMEGETTYPLILNIHGGPHMYGHTYFHEFQVLAAGKYAVVYINP 467
Db 358 SID-LELDGWTIKPEIKGEKA-PVIVFVHGGPKMGYGYEYKEMQLMASKRGYIVVYVNP 415
QY 468 RGSYGYGQEFVNAVGDYGGKDYDDVMAQVDAIKRDPHIDPKRLGVTTGGSGYGFMTNWI 527
Db 416 RGSNGYSEDFALRVLERTGLEDQFQILNGIEEFLRLEPOADRERIGITGISTGGTMTNWA 475
QY 528 VQOTNRKAAVTORSISNWIHFVGHVSDTYFTDQLEHDMFEDPEKLMWDRSPKLYAANV 587
Db 476 LTQSDLFRAKISENGISVWLSYAFSDIGLWF-DKEVIGDNPLENENYKLSPLFYAKNV 534
QY 588 EPTLILHGERDRDPIQEAOLFALKMKKQKTKLVFPNASHNLSTGHPQRQIKRLN 647
Db 535 KAPLLLIHSLDYRCPDQSLMFYHVLKDLGKREYVIAIFKGAHGHSIRGSPRRHMKRYK 594
QY 648 YISSWFDOHL 657
Db 595 LFMFEFFERKL 604

RESULT 10
B82580
alanlyl dipeptidyl peptidase XP2260 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-709 <SIM>
A:Cross-references: GB:AE004038; GB:AE003849; NID:g9107408; PIDN:AAF85059.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
J.D.; Junqueira, M.L.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
Chado, M.A.; Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
A:Authors: Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tshuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovskij-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:

A:Gene: XP2260
Query Match 13.6%; Score 473; DB 2; Length 709;
Best Local Similarity 22.7%; Pred. No. 8.2e-24;
Matches 163; Conservative 138; Mismatches 288; Indels 130; Gaps 23;
QY 9 DITAIIVSDTPOVAPDCGTRAAAYVKSOVNOEKDSTYSNIWIYETKTGGS-----VPWT 60
Db 49 DMVALDRVSSPELSPDGAVALVFAKQMDAKYIKASTSVWVQRLQAGTSAAPVRLTPLGMD 108
QY 61 HGEKRSTDPWSPDGRTLAFISDRGDAQAOLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db 109 -----VSAPVFSRDKAVYFLSAKSG-SHOLYVLPVSGGTSQLTNCAVDSDSKLSQ 162
QY 121 ESILVTISL-----CEGESIDDRKTEQDS-----YEPVEVQGLSKYKRDGKGLTRGAYA 169
Db 163 DRIVFSAVGFQVCGSDLSCTKRLKDEKNAKASGVVFEQLFVRHWDTWNDGRRNTL-FIA 221
QY 170 QLVIVSVK-----SGEMKELTSHKADHGDPAFSPDGKWLVSANITETDDA-S 216
Db 222 SLPAVGAKPVSVVSAMSAMLDGDPVKPFEGADHF--VMSPDGHSVVASIRVAGRQEPWS 279
QY 217 KPHDVIYIMLESQD-----LKQVTPHRGSGSSSPDGRYLLALLGNEKEYKNATLSKAWL 272
Db 280 TNFDLYRFDVSGHDAPVNLTAANPAWDA--TPMFSADGKMLYIRAMRPGFEADRFGLME 337
QY 273 YDIEOGRITCTLEMLDVHLADALIGSLGGAEQRPITWK-----DSQGFYVIGTD 323
Db 338 MEVQSGKV-----REIAPHWDRSADETALSADGKALYVNAAD 374
QY 324 QGSTGIYITSI-----EGLVYPIRLK---KEYINSFSLSPDEQHFIASTVTKPDRPS 371
Db 375 HGEHPLFKVDIAGSKVEKVGWEGSVHAPVLGAGKLAFAFARNLSKSAQDIETVDARVARG--- 431
QY 372 ELYSTPLGQEEKQLTGANDKFVREHTISIPETIQATYATEDGVVMVNGWLMRPAOMEGETTYP 431
Db 432 -----PL-----QAITSATGEVQLQVRLGDEQFSFKGWNDQETVYGVVWPKDYQPKKKY 482
QY 432 LILNIHGGPHMYGHTYFHEF--QVLAAGKYAVVYINPRGSHYGOEFVNAVGDYGGK 489
Db 483 VAFLIHGGPQSGFSGNSWGYRNPQTYAGGYAVVMIDFHGSGTGYGQAFDAISQHWGDRP 542
QY 490 YDDVMAQVDAIKRDPHIDPKRLGVTTGGSGYGFMTNWIIVGQTNRFRKAAVTQRSISNW--- 546
Db 543 LEDLQGWAAQAQOQYFPLNGDKACALGASVGGYGVYVYIAGHWNO-----PWKCL 591
QY 547 ISPHGVSD---IGY-----FFTDQLEHDMFEDTEKLMWDRSPKLYAANVETPLILHGER 598
Db 592 VOHDGVFDNRMMGYATEELWFSWENGSGTTPENPAGYEQFPNVLHVDKWRVPMVLVIHGQK 651
QY 599 DDCRPTQEAQELFALKMKKQKTKLVFPNASHNLSTGHPQRQIKRLNYSISSWFDOHL 657
Db 652 DFRIPVEQGLAFAFGALQKRGIESKLLYFDENHWLK---PQNSIQWHDIVNAWLKKYI 707
RESULT 11
H72474
probable acylamino-acid-releasing enzyme APE2441 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72474
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, Y.; Kudo, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: H72474
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-591 <KAW>
A:Cross-references: DDBJ:AP0000064; NID:g5105945; PIDN:BAA81456.1; PID:dl045242; PID:9
A:Experimental source: strain K1

A:Molecule type: mRNA
A:Residues: 1-732 <KOB>
A:Cross-references: GB:J04733; NID:g202931; PIDN:AAA8506.1; PID:g202932
A:Note: Parts of this sequence, including the amino end of the mature protein, were determined by Smith, J.A.
R:Lin, L.W.; Lee, F.J.S.; Smith, J.A.
Nucleic Acids Res. 17, 4397-4400, 1989
A:Title: Structural organization of the rat acyl-peptide hydrolase gene.
A:Reference number: S07624; MUID:89296508
A:Accession: S07624
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-732 <LIN>
A:Cross-references: EMBL:X14915
C:Genetics:
A:Introns: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 386/1
C:Superfamily: acylaminoacyl-peptidase
C:Keywords: blocked amino end; homotetramer; hydrolase; omega peptidase
F:1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental
F:118,291,443/Modified site: lysine derivative (Lys) #status experimental
F:587,675,707/Active site: Ser, Asp, His #status predicted

Query Match 11.4%; Score 396.5; DB 1; Length 732;
Best Local Similarity 21.7%; Pred. No. 1.1e-18;
Matches 148; Conservative 126; Mismatches 271; Indels 137; Gaps 21:

QY 22 APDGTAAVVK---SQVNOEKDSYTSNIW-----IYETKTGSGVPW 59
Db 97 SPSGTKAVLRKAGTGVSGEEKQL-EVWEKNRKLKSFNLSALEKHGHPVYEDDCFGCLSW 155
QY 60 THGEKRSTDPWSPDGRTLAFISDREGDAQAQLYIMSTEGGEARKLTDIPYGVSKPLWSPD 119
Db 156 SHSETH-----LLYVAEKK-----RPK 172
QY 120 GESILVTISLGEESIDDRKTEQDSYEPVEVQGLSKYKRDGKGLTRGAYLAQLVLVSKSG 179
Db 173 ASFFTKALDISASDEMARPCKPDQAIKQDFVYEDWGETMWSKSIPLVCLVDIDSG 232
QY 180 EMKEL--TSHKADHGDPAFSPDGKWLVSFA-----NLTTDDASKPHDVYIMSLSGDL 231
Db 233 NTSVLGVPENVSPGQAFWAPGDTGVVFGWHEPRLGTRYCTNRRSALYYVDLSGGK 292
QY 232 KQVTPHSGFSGSSFPDGRYLLALLGNEKEYKNATLSKAWLYDIEQRLTCLTEMLDVHL 291
Db 293 ELLSDGLAICSPLSPDQCRIVYLQVCLAPHQCSQLCYD---WTKVTSV-VDI 347
QY 292 ADALIGDSLIG-GAEQRP--WTKDS-----QGFYVIGTDQGSTGIYISIEG 336
Db 348 VPRQLGESFGIYCSLLPLGCSADSQRVVFDSQARSQDLFAVDITQGS--ITSUTAAG 405
QY 337 LVYPIRLEKEYINSFSLSPDEQHFASVTKPDRPSEL---YSIPLGOEE-----K 383
Db 406 SAGSWKL-----LTIDKLMVAQFSTPSPPLSKVGLFPLPPGKEQSVSWVSLSEAE 456
QY 384 QLTGANDKVFREHTISPEEIQYATEDGVVNGWMLMPAQMEGETTYPILNIHGGPHM 443
Db 457 PIPGIHGVRLVLPDPQENQYADLDFEAL---LLQPSNPDKTQVPMVMPHGGPHSS 513
QY 444 YGHTYFEHFOVLAAGYAVVYNPRGSHGYGQEFVNAVVRGDKYDDVNMQAVDEAIKR 503
Db 514 FVTAMFLFAPMLCKMGFAVLLVNRGTFGQDSILSLPGNVGHQDKVDQFAVEQVLOE 573
QY 504 DPHIDPKRLGVGGSGYGFMTNIWVGQ--TNRFKAATVQTSISNWSIFHGVSDIGYFTDW 562
Db 574 E-HFDARRVALMGSGHGLSCHLIGQYPTYSACIARNPVNTIASMMGSTDI-----PDW 628
QY 563 QLEHDMFEDT-----EKLWDRSPLKYAANVETPLLLIHGRRDRRCPTQAEQLEFI 612
Db 629 CMVETGFFYSNCLPDNLNWEMLDKSPIKIPQVKTTPVLLMLGQEDRRVFPKGMYYR 688
QY 613 ALKKMGKETKLVRFNPNASHNLS 634
Db 689 ALKARNVVPVRLVLLYPKSNHLS 710

RESULT 14
J00132
N:Alternate names: acyl-peptide hydrolase; acylamino-acid-releasing enzyme
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 21-Jul-2000
C:Accession: J00132
R:Mitta, M.; Asada, K.; Uchimura, Y.; Kimizuka, F.; Kato, I.; Sakiyama, F.; Tsunasawa
J. Biochem. 106, 548-551, 1989
A:Title: The primary structure of porcine liver acylamino acid-releasing enzyme deduc
A:Reference number: J00132; MUID:90110044
A:Accession: J00132
A:Molecule type: mRNA
A:Residues: 1-732 <MIT>
C:Genetics:
A:Introns: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 3
C:Superfamily: acylaminoacyl-peptidase
C:Keywords: acetylated amino end; homotetramer; hydrolase; omega peptidase
F:1/Modified site: acetylated amino end (Met) #status experimental
F:587,675,707/Active site: Ser, Asp, His #status predicted

Query Match 11.2%; Score 391; DB 1; Length 732;
Best Local Similarity 21.0%; Pred. No. 2.7e-18;
Matches 149; Conservative 123; Mismatches 287; Indels 152; Gaps 23:

QY 22 APDGTAAVVKSQ--VNOEKDSYTSNIW-----IYETKTGSGVPWT 60
Db 97 SPSGTKAVLRKAGTGTAEKQFLEWKNRKLKSFNLSALEKHGHPVYEDDCFGCLSW 156
QY 61 HGEKRSTDPWSPDGRTLAFISDREGDAQAQLYIMSTEGGEARKLTDIPYGVSKPLWSPDG 120
Db 157 HSETH-----LLYVAEKK-----RPKA 173
QY 121 ESTLVITISLGEESIDDRKTEQDSYEPVEVQGLSKYKRDGKGLTRGAYLAQLVLVSKSGE 180
Db 174 ESFFTKALDVTGSDDEMARTKPDQAIKQDFLEWGENMVSKSTPVLCLVDIESGN 233
QY 181 MKEL--TSHKADHGDPAFSPDGKWLVSFA-----NLTTDDASKPHDVYIMSLSGDLK 232
Db 234 ISVLEGVPESVSPQAFWAPGDTGVVFGWHEPRLGTRYCTNRRSALYYVDLTGKCE 293
QY 233 QVTPHSGFSGSSFPDGRYLLALLGNEKEYKNATLSKAWLYDIEQRLTCLTEMLDVHLA 292
Db 294 LLSDSVAVTSPRLSPDQCRIVYLRFPFLVPHQCGQLCYD---WTKVTSV-VDI 348
QY 293 DALIGDSLIG-GAEQRP--WTKDS-----QGFYVIGTDQGSTGIYISIEGL 337
Db 349 PRQLGEDFGIYCSLLPLGCSADSQRVVFDSQARSQDLFAVDITQGSVTSLTAGSGG 408
QY 338 VYPIRLEKEYINSFSLSPDEQHFASVTKPDRPSEL---YSIPLGOEEKQITGANDKFEVR 394
Db 409 SWKL-----LTIDRLMVVQFSTPSPPLSKVGLFPLPAGKEQ-----AVSWVSL 452
QY 395 EHTISIP-----EETQYATEDGVVNGWMLMPAQMEGETTYPILNIHGG 439
Db 453 BEAEFPDIDSIRVLQPPQOEHVQYA---GLDPEALLOQPSNPSPEKTQVPMVMPHGG 509
QY 440 PHMYGHTYFIEFQVLAAGYAVVYNPRGSHGYGQEFVNAVVRGDKYDDVNMQAVDE 499
Db 510 PHSSFVTAMFLFAPMLCKMGFAVLLVNRGTFGQDSILSLPGNVGHQDKVDQFAVEQ 569
QY 500 AIKRPDHPIDPKRLGVGGSGYGFMTNIWVGQ--TNRFKAATVQTSISNWSIFHGVSDIGYF 558
Db 570 VLQEE-HFDARRVALMGSGHGLSCHLIGQYPTYSACVVRNPVNTIASMMGSTDI----- 625
QY 559 FTDWOLEHDMFEDTE-----KLW-----DRSPLKYAANVETPLLLIHGRRDRRCPTQAE 608
Db 626 -PDWCMVEAGFSYSSDCPLDLSVWAAMLDKSPIKIPQVKTTPVLLMLGQEDRRVFPKQGM 684

QY 609 QLFIALKKMGKTKLVRFNASHNLSTGHPRQRIKRLNYISS--WFDQHL 657
 Db 685 EYRVLRKARVVPVRLILYPKSTHALSEV-----EVESDSFMNAVLMCLTHL 730

RESULT 15
 JC4655
 acylaminoacyl-peptidase (EC 3.4.19.1) - human
 N:Alternate names: acidpeptide hydrolase; acylamino acid releasing enzyme; lung DNF15S2
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: JC4655; A42257; A30145; A53799
 R:Witt, M.; Ohnogi, H.; Mizutani, S.; Sakiyama, F.; Kato, I.; Tsunasawa, S.
 DNA Res. 3, 31-35, 1996
 A:Title: The nucleotide sequence of human acylamino acid-releasing enzyme.
 A:Reference number: JC4655; MUID:96281126
 A:Accession: JC4655
 A:Molecule type: mRNA
 A:Residues: 1-732 <M17>
 A:Cross-references: DDBJ:D38441; NID:g556513; PIDN:BAA07476.1; PID:g556514
 A:Experimental source: liver
 R:Scaloni, A.; Jones, W.M.; Barra, D.; Pospischil, M.; Sassa, S.; Popowicz, A.; Manning,
 J. Biol. Chem. 267, 3811-3818, 1992
 A:Title: Acylpeptide hydrolase: inhibitors and some active site residues of the human en
 A:Reference number: A42257; MUID:92156118
 A:Accession: A42257
 A:Molecule type: protein
 A:Residues: 418-453; 90-95; 119-126; 130-136, 'A', 138-149; 173-193, 'T', 195-199, 'A', 201-219
 71, 'O', 473-475, 'V', 477-497; 519-527; 539-574, 'M', 576-591, 'L', 593; 656-657, 'K', 659-664; 677-6
 A:Experimental source: lung
 R:Naylor, S.L.; Marshall, A.; Hensel, C.; Martinez, P.F.; Holley, B.; Sakaguchi, A.Y.
 Genomics 4, 355-361, 1989
 A:Title: The DNF15S2 locus at 3p21 is transcribed in normal lung and small cell lung can
 A:Reference number: A30145; MUID:89233127
 A:Accession: A30145
 A:Molecule type: mRNA
 A:Residues: 102-112, 'D', 114-136, 'A', 138-167, 'K', 169-196, 'QTPRSIGISCFMKTRGNMYSKSIPIVSACW
 ALQIRPKCPWMSCT', 509, 'A', 511-581, 'X', 583-604, 'RACURTRDHNASMLGSTDLTGAWW', 'RL', 617-618, 63
 A:Cross-references: GB:J03068; NID:g181629; PIDN:AAA35769.1; PID:g181630
 A:Experimental source: lung
 R:Scaloni, A.; Barra, D.; Jones, W.M.; Manning, J.M.
 J. Biol. Chem. 269, 15076-15084, 1994
 A:Title: Human acylpeptide hydrolase. Studies on its thiol groups and mechanism of actio
 A:Reference number: A53799; MUID:94253066
 A:Accession: A53799
 A:Molecule type: protein
 A:Residues: 19-33; 63-65 <SC2>
 C:Genetics:
 A:Gene: GDB:APDH; AARE: D3s48E; DNF15S2
 A:Cross-references: GDB:127917; OMIM:102645
 A:Map position: 3p21.3-3p21.2
 C:Function:
 A:Description: hydrolyzes amino-terminal acetylated peptides into peptides having free a
 C:Superfamily: acylaminoacyl-peptidase
 C:Keywords: acetylated amino end; blocked amino end; homotetramer; hydrolase; liver; ome
 F:1/Modified site: acetylated amino end (Wet) #status predicted
 F:58/Active site: Ser #status experimental
 F:675,707/Active site: Asp, His #status predicted

Query Match
 Best Local Similarity 11.0%; Score 384; DB 1; Length 732;
 Matches 147; Conservative 136; Mismatches 287; Indels 130; Gaps 24;

QY 22 APDGTAAAYVRS-----QVNCXKSYSTNIW-----IYETKTGGSVPWT 60
 Db 97 SPGSMKAVLRKAGTGGEKEQFLVWEKNRKLKSFNLVLEKHGVPVYEDDCFGCLSW 156
 QY 61 HGEKSTDPKRWSPGRTIAFISDREGDAALYINSTEGEARKLTDIPYGVSKPLWSPDG 120
 Db 157 HSETH-----LLYVAERK-----RPKA 173

QY 121 ESILVITISLGGESIDDRKTEQDSYEPVE-VQGLSYKRDGKGLTRGAYAOVLVLVSKSG 179
 Db 174 ESFTQKALDVASASDDEIARLKPD-QPIKGDQFVYEDWGENMYSKSIPLVLCVLDVSSG 232
 QY 180 EMKEL--TSHKADHGDPAESPDGKWLVSFA-----NLTETDDASKPHDVYIMSLSGDL 231
 Db 233 NISVLEGVPENVSPQAFWAPGDAGVYVGVGMHEPFRGLGTRFCTNRRSALYVVDLIGGKC 292
 QY 232 KQVTPHRGSGSSFSFDPGRYLLALLGKEYKNATLSKAWLYDIEQGRLLCTCLTEMLDVHL 291
 Db 293 ELLSDSLAVSSPRLSPDQCRIVVLOVPSLPHHQCQLCLYD----WYTKVTSVV-VDV 347
 QY 292 ADALIGDSLIG-CAEQRPV--WTKDSOGFYVIGTDQSGTGIIYISIE-GLVYPIRLKEY 347
 Db 348 VPQLGENFSYICSLPLPCGWSADSORVVVDFSAQSRQDLFAVDTOGVTVTSLTAGSG 407
 QY 348 INSFSLSPDEQHFIAVTKPDRFELY--SIPLGQEEKQLTGANDKFVREHTISIP---- 401
 Db 408 GSKMLLTIDQDLMAVOFSTPLPPTLKVGFLPSAGKEQSVLWVS-----LEEAPIPDIHW 463
 QY 402 -----PEIOXATEDGVNMGWLMRPAQMEGETTYPLINIHGGPHMYGHTYFH 450
 Db 464 GIRVLQPPPEQENVQTA---GLDFAILLQPGSPDKTQVPMVMPHGGPHSFSFVTAWML 520
 QY 451 EFQVLAAGYAVVYINPRGSHGYGQEFVNAVGRDYGKDYDDVMQAVDEAKRDHPIDPK 510
 Db 521 FPAMLCCKMGFAVLLVNYRGSTGFGQDSIILPGNVGHQDKVDQFAVEQVLOEE-HFDAS 579
 QY 511 RLGVTVGGSGGPMTWIVGQ--TNRKAAVTRQSIISNWSIFHGVSDFGYEFTQW----- 562
 Db 580 HVALMGSGHGGFISCHLLIGQYETRYACVARNPVINIASMLGSTDI----PDMCVVEAGF 635
 QY 563 QLEHDMFEDTE---KLWDRSPLKYAANVETPLLLHGERDDRCPIEQABQLFIALKKMGK 619
 Db 636 PFSSDCLPDLSSWAEMLDKSPIRYTPQVKTPLLMLGOEDRRVPPKQGMYYRALKTRNV 695
 QY 620 ETKLVRFNASHNLSTGHPRQRIKRLNYISS--WFDQHL 657
 Db 696 PVRLILYPKSTHALSEV-----EVESDSFMNAVLMCLTHL 730

Search completed: September 25, 2001, 17:37:39
 Job time: 146 sec

us-09-462-845-2.rpr

Wed Sep 26 11:29:29 2001

[illegible]

157	Db				-----LLYVAERK-----	RPKA	179
121	2y	ESILVTTISL	GESIDREKTEQDSYEPVE	-VQGLSYKRDGKGLTGAYAOVLVSVKSG			179
174	Db	ESFQTIKALD	VSADDEIARLKKPD-QPIKQDQFVEDNGENMVSKSIPVLCVLDES				232
180	2y	EMKEL--TSHKADH	GPAPSPDGKWLVPFA-----NUTETDDAKPHDVIYMSLES	GD			231
233	Db	NISVLGV	PENVSPGOAFWAPCDAGVFPVGMWHIEPFRILGIRFCTNRRSALYVDVLIGKC				292
232	2y	KQVTPHRC	SFGSSSPDGRYLAALLGNEKEYKNATLSKAWLYDTEQGRLTCLTEMLDVHL				291
293	Db	ELLSDDSL	AVSPRLSPDQCRIVLYQESLPHHOC	SQCLYD---WTKVTSV-VDV			347
292	2y	ADALIGDSLIG	-GAEORPI--WTKDSQGFYVIGTDQGSTGIYIYSIE	-GLVYPIRLEKEY			347
348	Db	VPRLGENF	SGIYCSLLPLGWSADSRQVFD	SAQRSQDLFAVDVQVGTVTSLTAGSG			407
348	2y	INFSLS	DPDEQHFIVASVTKPDRPSELY--SPLGEEKOLTGANDKFVREHTISIP	----			401
408	Db	GSWKLLT	IDQDLMAVQFSPSPPTLKVGLFSA	CKESVLWS---LEEAEPID	IHW		463
402	2y	-----EEIQ	ATBDGVYVNGWLMRPAQMEGETTYPLILNHGSPHMYGHTYFH				450
464	Db	GIRVLQPP	PEQENVQYA---GLDEFAILLQPGSPDKTQVPMVMPHGPHSFSFTAWML				520
451	2y	EFQVLA	KAYAVVYINPRGSHGYGQEFVNVRG	DYGGKDYDDVMOAVDEAIKRDPHIDPK			510
521	Db	FPAMLCK	MGFAVLLVNYRGSTGFGDSILSPGNVGHQVDKQVAVEGVLQEE	-HFDAS			579
511	2y	RLGVTC	SGYGGFTWNIWVGQ--TNRFKAAVTQRSINWISFHGVS	DIGYFTFW-----			562
580	Db	HVALMG	SGHGGLFISCHLIGQYETRYACVARNP	VINIASMLGSTD	----PDWCVV	EAGF	635
563	2y	QLEHDM	FEDETE---KLWDRSP	KYKANVETPLLILHGERDDRCPTQASQLFALKMKMG			619
636	Db	PSSDC	LPDLVSVAEMDLSPIRYIPQVTF	LLLLLMLGQEDRRVPFKQGM	EYVRA	LKATRN	695
620	2y	ETKLIV	FPNASHNLSTRGHPQRIKLN	YISS--WFDQHL			657
696	Db	PVRLIL	PKSTHALSEV-----EVESDS	SPNAVILWLRTHL			730
RESULT	5						
YL31_CAEEL		STANDARD;		PRT;		761	AA.
AC		P34422;					
DT		01-FEB-1994		(Rel. 28, Created)			
DT		01-FEB-1994		(Rel. 28, Last sequence update)			
DE		30-MAY-2000		(Rel. 39, Last annotation update)			
DT		HYPOTHEtical		86.0 KDA PROTEIN F44B9.1 IN CHROMOSOME III.			
GN		F44B9.1.					
OS		Caenorhabditis elegans.					
OC		Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;					
OC		Rhabditidae; Peloderinae; Caenorhabditis.					
NCBI_TaxID=		6239;					
[1]		SEQUENCE FROM N.A.					
RC		STRAIN-BRISTOL N2;					
RC		MEDLINE=94150718; PubMed=7906398;					
RA		Wilson K., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R., Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Watson R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;					
RA		"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.					

DR InterPro: IPR002471; .
DR Pfam: PF00930; DPIP_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO-ENDOPEP_SBR; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 766 MEMBRANE FORM OF DPP (MDPP).
FT CHAIN 39 766 SOLUBLE FORM OF DPP (SDPP)
FT (BY SIMILARITY).
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT TRANSMEM 1 6
FT DOMAIN 7 28
FT DOMAIN 29 766
FT ACT_SITE 630 630
FT ACT_SITE 708 708
FT ACT_SITE 740 740
FT CARBOHYD 85 85
FT CARBOHYD 92 92
FT CARBOHYD 150 150
FT CARBOHYD 219 219
FT CARBOHYD 229 229
FT CARBOHYD 281 281
FT CARBOHYD 321 321
FT CARBOHYD 520 520
FT CARBOHYD 685 685
FT CONFLICT 6 6
FT CONFLICT 7 7
FT CONFLICT 437 437
FT CONFLICT 557 557
FT CONFLICT 563 563
FT CONFLICT 663 663
SQ SEQUENCE 766 AA: 88278 MW; 5FB42C6662D6117 CRC64;

Query Match 7.8%; Score 272; DB 1; Length 766;
Best Local Similarity 20.4%; Pred. No. 3 le-10;
Matches 153; Conservative 113; Mismatches 285; Indels 200; Gaps 34;
QY 15 SVTDPOYADPDGRAAVKQVQKDSYTSNIWIY-----EYKGGVPTWTHGEKR 65
DB 101 SINDYISPDGQFILLEYNVVKWRHSYASYDIYDLNKRQLITEERIPNNTOMV----- 155
QY 66 STDPWSPDGRFLAFISDREGDAALYINSTEGERAKLT-----DIPY----- 109
DB 156 ---TWSPVGHKLAYVNNND-----IYVKEPNLPYRIWTGKEDIYNGITDWYEE 206
QY 110 ---GVSKPLWSPDGESILVTISLGEISIDRE--KTEQDSYEFVEVQGLSKYRDGKGLT 164
DB 207 VFSAYSALWSPNGTFLAYA-----QFNDTEVPLIEYSFYSDLSQ----- 247
QY 165 RGAYAOVLVSV--KSGEMKELTSHKADHGDPAFSPDGKWLVSANLTETDDASKPHDVI 223
DB 248 ---YPKTVRVPYPKAG-----AVNPTVKFFV-----VNTDLS----- 277
QY 224 MSLESGDLKQVTPHRGSGSFSFSPDGRYLALLGNEKEYKNAT---LSKAWL----- 272
DB 278 -SVTNATSIQT-----APASMLIGDHYLC-----DVTWATQERISLOWLRIONYSV 324
QY 273 -----YDIEQGRFLCTEMLDVHLADALIGDSLIGG-AEORPIWTKDSOGFY-VIGTDQG 325
DB 325 MDICDYDESSGRWNCVLVARQHIMSTT-----GWVGRFPRSPSEPHFTLDGNSFYKIIISNEG 380
QY 326 STGIYIISIE--GLVYPIRLEKEYINSFSLSPDQHQHFIATSVK----- 366
DB 381 YRHICYFDKKDCDTFTKGTWETVIGIEALTSYLYIISNEYKMPGGRNLYKIQLSDYT 440
QY 367 -----PDRPSELYSIFLQGEK--QL-----TGANDKFFVR--EHTIS 399
DB 441 KVTCLSCENLPER-CQYYSVSFSKEAKYIQLKSCGPGFLYTLTHSSVNDKGLRVLEDNSA 499
QY 400 IPEIQYATE-----DGVMVNG---W---LMRPAQMEGETTYPLIINIHGGPHMYGHYTF 449
DB 500 LDKMLQNVQMPFSKKLDFIILNETKEFYQMLPPHFDKSKYPLLDVYAGPCSKADTVF 559
QY 450 H---EFQVLAARGYAVVYINPRGSHGYGOEFVNAVGRDYGKDYDDVMQAVDEAIKRDHP 506

RX MEDLINE-96067599; PubMed-7487939;
RA Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT regulation from a TATA-less GC-rich sequence characteristic of a
RT housekeeping gene promoter.";
RL Biochem. J. 311:835-843(1995).
RN [8].
RP PARTIAL SEQUENCE.
RC TISSUE-Kidney;
RX MEDLINE-93210468; PubMed-8096237;
RA Morrison M.E., Vijayaradhhi S., Engelstein D., Albino A.P.,
RA Houghton A.N.;
RT "A marker for neoplastic progression of human melanocytes is a cell
RT surface ectopeptidase.";
RL J. Exp. Med. 177:1135-1143(1993).
CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC PENULTIMATE RESIDUE IS PROLINE.
CC -1- CATALYTIC ACTIVITY: DIPEPTIDYL-POLYPEPTIDE + H(2)O = DIPEPTIDE +
CC POLYPEPTIDE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC A SOLUBLE FORM.
CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC BY PROTEOLYTIC PROCESSING.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC PROLYL OLIGOPEPTIDASE FAMILY.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U13735; AAB60646.1; JOINED.
DR EMBL: U13710; AAB60646.1; JOINED.
DR EMBL: U13711; AAB60646.1; JOINED.
DR EMBL: U13712; AAB60646.1; JOINED.
DR EMBL: U13713; AAB60646.1; JOINED.
DR EMBL: U13714; AAB60646.1; JOINED.
DR EMBL: U13715; AAB60646.1; JOINED.
DR EMBL: U13716; AAB60646.1; JOINED.
DR EMBL: U13717; AAB60646.1; JOINED.
DR EMBL: U13718; AAB60646.1; JOINED.
DR EMBL: U13719; AAB60646.1; JOINED.
DR EMBL: U13720; AAB60646.1; JOINED.
DR EMBL: U13721; AAB60646.1; JOINED.
DR EMBL: U13722; AAB60646.1; JOINED.
DR EMBL: U13723; AAB60646.1; JOINED.
DR EMBL: U13724; AAB60646.1; JOINED.
DR EMBL: U13725; AAB60646.1; JOINED.
DR EMBL: U13726; AAB60646.1; JOINED.
DR EMBL: U13727; AAB60646.1; JOINED.
DR EMBL: U13728; AAB60646.1; JOINED.
DR EMBL: U13729; AAB60646.1; JOINED.
DR EMBL: U13730; AAB60646.1; JOINED.
DR EMBL: U13731; AAB60646.1; JOINED.
DR EMBL: U13732; AAB60646.1; JOINED.
DR EMBL: U13733; AAB60646.1; JOINED.
DR EMBL: U13734; AAB60646.1; JOINED.
DR EMBL: M74777; AAA51943.1; .
DR EMBL: M80536; AAA52308.1; .
DR EMBL: X60708; CAA43118.1; .
DR EMBL: S79876; AAB35614.1; .
DR PIR: S24313; CDHU26.
DR MIM: 102720; .
DR InterPro: IPR001375; .
DR InterPro: IPR002469; .

Db 560 RLNWATYLASTENIIVASFDGRSGYQDKIMHAINRLRGTEVEDEQIEAARQFSKMG-F 618
 QY 507 IDPKRLGVGGSGGFWTWIYVQ--TNRKFAAVTORSISNWSIFHGVSDIGYFFTDWQLE 565
 Db 619 VNKRIAIWGSVGGVYVTSWVLGSGVFKGAGIAPVSRWEYDSV-----YTERYMG 672
 QY 566 HDMFEDTEKILDRSPKLYAAN--VETPLILHGERDRCPICQAEQLFTALKMKCKETKL 623
 Db 673 LPTPEDLDHNRSTVMSRAENFKQVEYLLINGTADNNVHFOOQAISKALVDVGVDFQA 732
 QY 624 VRFPNASHNL-SRTHCPQRIRKRLNYISSWF 653
 Db 733 MWYTTDEHGIIASSTAHOHYTHMSHFIRKQCF 763

RESULT 7

DAPI_YEAST
 ID DAPI_YEAST STANDARD; PRT; 931 AA.
 AC P33894;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DIPEPTIDYL AMINOPEPTIDASE A (EC 3.4.14.-) (DPAP A) (YSCIV).
 GN STE13 OR YC11 OR YOR219C OR YOR50-9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95066382; PubMed=7975897;
 RA Anna-Arriola S.-S., Herskowitz I.;
 RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
 RL aminopeptidase.";
 RL Yeast 10:801-810(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB320;
 RA Flanagan C.A., Thorne J.;
 RT "STE13.";
 RL (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
 RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
 RN Press, Oxford (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96437977; PubMed=8840505;
 RA Gallissou F., Dujon B.;
 RT "Sequence and analysis of a 33 kb fragment from the right arm of
 RL chromosome XV of the yeast Saccharomyces cerevisiae.";
 RL Yeast 12:877-885(1996).
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
 CC ALPHA-FACTOR PRECURSOR.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
 CC VACUOLES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO DRAP B.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcements>
 or send an email to license@isb-sib.ch).

EMBL; L21944; AAA35119.1;
 EMBL; U08230; AAA17897.1;
 EMBL; X92441; CAA63182.1;
 EMBL; Z75127; CAA99437.1;
 PIR; A49737; A49737;
 PIR; S45451; S45451.

DR MEROPS: S09.005;
 DR SGD: S0005745; STE13.
 DR InterPro: IPR001375;
 DR InterPro: IPR002469;
 DR InterPro: IPR002471;
 DR Pfam: PF00930; DPIP_N_term; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: P500708; PRO_ENDOPEP_SER; FALSE_NEG.
 DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
 FT DOMAIN 1 119
 FT TRANSMEM 120 140
 FT DOMAIN 141 931
 FT ACT_SITE 785 785
 FT ACT_SITE 863 863
 FT ACT_SITE 896 896
 FT CARBOHYD 377 377
 FT CARBOHYD 814 814
 SQ SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

Query Match

Best Local Similarity 7.7%; Score 269; DB 1; Length 931;
 Matches 155; Conservative 108; Mismatches 298; Indels 212; Gaps 30;

QY 9 DITAIIVTDPQYAPDGTAAAYKVSQVNOEKDS--YTSNI-----WIYETKG 54
 Db 220 DTFEVLNIGNRFLYEGVEFTVSTQVINYKLDKILFGTNLESEFHSKGFYWKDLNTG 279
 QY 55 GSVPTWTHCEKSTYD-----PRWSPDGRTLAFISDRGDAALYIMSTEGEARKL 104
 Db 280 NIEPILPEKSDNNYELGSLKSLYAHSFPAINYFYFVEN-----NLFQOVNSGVAKKV 334
 QY 105 TDIPYGVSKPL-----WSPDGSILVTISLGSSEIDDRKTEQ 143
 Db 335 TE--DGSKDIFNAKPDWYEEVLAADQAIWAPDDSK--AVTARFNDTSVDDILRLNY 389
 QY 144 DSYEPVEVGLSYKRDGKGLTRGAYAOILVSVKSGEMKEL-TSHKADHGDPAPSPDGKW 202
 Db 390 TNMEAYLSDTKIKYKPGFQNPQF-DLFLVNLQNGIYISINTGGQD---SILYNGKW 444
 QY 203 L---VFSANLTTEDDASKPHDVIYIMSLESG-----DLKQVTP-- 236
 Db 445 ISPDTRFEI--TDNRSKILDVYDIPSSOMLTVRTWNSNLFNGWIEKTKDILSIPKP 502
 QY 237 --HRGSGSSSPDGRYLALL-----GNEKEYKNATLSKAWLYDIEQG 278
 Db 503 ELKRMIDYGYIDIHADSRGSHLFYPTVFAKEPIQLTKGNWEVTGNGIVG---YEYTD 558
 QY 279 RLTCLEMLDV---HLADALIGSLIGGAEQRPINWKDSQGVYVIGTDQSGTGIYISIE 335
 Db 559 TIFTTANEICVMSQHLYSISLSTDTTONTFQSLQNFSDKYDFY----- 601
 QY 336 GLVYPIRLKEYINFSLSFDEQHFIASTKPDPRP-----SELYSTIPGOEE 382
 Db 602 -----DFELSSARYAISKLLGPDTPIKVAGPLTRVLNVAIHDDLSILQLT 647
 QY 383 KQLTGANDKF--VREHTTISIPEEIYATYEDGVVWNGWLMRPAOMGETTYPLILNHGG 439
 Db 648 K-----DEAFKEKIKYNDIPITYKTMVLDDGVGEINYEIKPANLNPKKKYPLVNIYGG 702
 QY 440 PHMYGHTYF-----HEFQVLAAGYAVVYINPRGSHGYGOEFVNAVYGDYGGKDYDD 492
 Db 703 P-----GSQFTTKSSLAFFQAVVSGLDVILQLEPRCTGCKGWSFKSWAREKLGWYEPDR 758
 QY 493 VMOAVDEALKRD-PHIDPKRLVGTSGSGFMTNIVGQTN--RFAKAAVTSRISNWSIF 549
 Db 759 ITEVTKFKFIORSQHIDESKIAIWGVSYGFTSLKTVELDNGDTFKYAMAVAVPNTWLY 818
 QY 550 HGV-----SDIGYFFTDWQLEHDMFEDTEKLDNRSPKLYAANVETPLILHGER 598
 Db 819 DSVYTERYMNQPSNHEGYFEVSTIONPKSFESKRLF-----IVHGTF 862

Wed Sep 26 11:29:29 2001

us-09-462-845-2.rsp

QY 599 DDCRPIEQAEOLFALKMG-KETKLVRFNPASHNHSRTGHPQRI--KRLNY 648
 DB 863 DDVHIONTRFLVDQLNLLGLTYNDHIFPDSHSIRY--HNAQRIVFQKLYY 913

RESULT 8
 ID PP4_MOUSE STANDARD; PRT; 760 AA.
 AC P28843; 1992 (Rel. 24, Created)
 DT 01-DEC-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 40, Last annotation update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5) (DPP IV) (THYMOCYTE-ACTIVATING
 MOLECULE) (THAM).
 GN DPP4

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

[1]
 R1 SEQUENCE FROM N.A.
 R2 STRAIN=SWISS; TISSUE=Thymus;
 RX MEDLINE=92129288; PubMed=1370813;
 RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
 RA Pierres M.;
 RA "cDNA cloning for mouse thymocyte-activating molecule. A
 RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
 RT subgroup of serine proteases.";
 RL J. Biol. Chem. 267:2200-2208(1992).
 RN [2]

RP REVISIONS.
 RA Marguet D.A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=B10.A; TISSUE=Liver;
 RC MEDLINE=9302780; PubMed=7999781;
 RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
 RA "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
 RL Biochemistry 33:15204-15214(1994).
 RN [4]

RN SEQUENCE OF 1-20.
 RX MEDLINE=91302787; PubMed=1712807;
 RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
 RA Bernard A.-M., Gorvel J.-P., Pierres M.;
 RA "Evidence that thymocyte-activating molecule is mouse CD26
 RT (dipeptidyl peptidase IV).";
 RL J. Immunol. 147:447-454(1991).
 CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
 CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
 CC PENULTIMATE RESIDUE IS PROLINE.
 CC -1- CATALYTIC ACTIVITY: DIPEPTIDYL-POLYPEPTIDE + H(2)O = DIPEPTIDE +
 CC POLYPEPTIDE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
 CC A SOLUBLE FORM.
 CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
 CC BY PROTEOLYTIC PROCESSING.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X58384; CAA41274.1; -
 CC EMBL; U12620; AAA82213.1; -
 CC EMBL; U12599; AAA82213.1; JOINED.
 CC EMBL; U12600; AAA82213.1; JOINED.

DR EMBL; U12601; AAA82213.1; JOINED.
 DR EMBL; U12602; AAA82213.1; JOINED.
 DR EMBL; U12603; AAA82213.1; JOINED.
 DR EMBL; U12604; AAA82213.1; JOINED.
 DR EMBL; U12605; AAA82213.1; JOINED.
 DR EMBL; U12606; AAA82213.1; JOINED.
 DR EMBL; U12607; AAA82213.1; JOINED.
 DR EMBL; U12608; AAA82213.1; JOINED.
 DR EMBL; U12609; AAA82213.1; JOINED.
 DR EMBL; U12610; AAA82213.1; JOINED.
 DR EMBL; U12611; AAA82213.1; JOINED.
 DR EMBL; U12612; AAA82213.1; JOINED.
 DR EMBL; U12613; AAA82213.1; JOINED.
 DR EMBL; U12614; AAA82213.1; JOINED.
 DR EMBL; U12615; AAA82213.1; JOINED.
 DR EMBL; U12616; AAA82213.1; JOINED.
 DR EMBL; U12617; AAA82213.1; JOINED.
 DR EMBL; U12618; AAA82213.1; JOINED.
 DR EMBL; U12619; AAA82213.1; JOINED.
 DR PIR; S23752; S23752.
 DR MEROPS; S09.003; -
 DR MGD; MGI:94919; Dpp4.
 DR InterPro; IPR001375; -
 DR InterPro; IPR002469; -
 DR InterPro; IPR002471; -
 DR Pfam; PF00930; DPP4_N-term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 DR Hydroxylase; Amino-peptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 760
 FT CHAIN 37 760
 FT DOMAIN 1 6
 FT TRANSMEM 7 28
 FT DOMAIN 29 760
 FT DOMAIN 295 466
 FT ACT_SITE 624 624
 FT ACT_SITE 702 702
 FT ACT_SITE 734 734
 FT CARBOHYD 83 83
 FT CARBOHYD 90 90
 FT CARBOHYD 113 113
 FT CARBOHYD 213 213
 FT CARBOHYD 223 223
 FT CARBOHYD 315 315
 FT CARBOHYD 328 328
 FT CARBOHYD 514 514
 FT CARBOHYD 679 679
 FT SEQUENCE 760 AA; 87436 MW; ASF644B46A43DF8 CRC64;

Query Match 7.1%; Score 247; DB 1; Length 760;

Best Local Similarity 19.9%; Pred. No. 1.3e-08;

Matches 145; Conservative 126; Mismatches 281; Indels 178; Gaps 34;

QY 2 KKLITADDI---TAIVSVTDPOYAPDGTAA-----YVKSOVNOEKDSVTS---NI- 46
 DB 134 RQLITEKIPNNTQWIT-----WSPEGHKLAYVWKNDIYVKVEPHLSHRITSGEENVI 188
 QY 47 -----WIYETKGG--SVPTWTHGEKRSRSTDPWSPDGRITLAF--ISDREGDAAQLYIMST 96
 DB 189 YNGITDWYEEVEFGAYSALW-----WSPNNTFLAYAQFNDTGVPLIEVSFYS 237
 QY 97 EGGEARKLITDIPYGVSKPLMSPDGESILVTISLGESESIDREKTEQDSYEPVEV----- 151
 DB 238 ESLQYPTKVTWIPYFKAGAV-NPTVKFFIVNID-----SLSSSSAAPIQIPAPAS 286
 QY 152 --QGLSKYRDKGKGLTRGAYAQVLVSVKSGEMKELTSHKADHGDPAFSPDGLVFSANL 209
 DB 287 VARGDHYLCVDVWVWATEERISLQWLRIQNIYSVMAICDY-----KINL 329
 QY 210 TETDDASKPHDVYIMSLSGDLKQVTPHRGSGFGSSSF-----SPDGRYLLALLGNEKEYKNA 265

DR PIR; A33315; A33315.
DR MEROPS; S09.003; -
DR InterPro; IPR001375; -

QY 15 SVTDPQYADCTRAAYVKSQVNKEDSYTSNIWY-----ETKTGSGVPHTHCEKR 65
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Matches 134; Conservative 122; Mismatches 283; Indels 197; Gaps 26;
Local Similarity 16.24; Pred. No. 2.5e-08;

Query	Match	Best Local Similarity	7.0%; 18.2%; Matches 134;	Score 242.5; Pred. No. 2.5e-08; Conservative 122;	Mismatches 283;	Indels 197;	Gaps 26;
15	SVTDPDYADGCTRAAYKVSQVNOEKSYTSINIIY		-----	-----	-----	-----	-----
99	SISDYSVSPDLVLEYNKVMRHSYTSYSLYDLNKRQLITEKIPNNQWII		-----	-----	-----	-----	-----
db							


```

QY 335 -----EGLVPIRLKEKYNISLSLSPDQHQHTASVTKDPRPSELY 374
Db 463 QPNSSNDNIOSITSGMDVTEILTYDEKRNKLYFLSTEDLPRRHLYSANTVDDFNQCL 522
QY 375 SIBL-----QOEKQLTGANDKFKVREHTISIPETIOYATE 409
Db 523 SCDLVENCVTVSASFHNDFLLKCEGCPVPTVTHNTDK-RMFDLEANEQVOKAIY 581
QY 410 DGVN-----VNGW-----LMRPAQMEGETTYPLILNIHGPHMYGHTYFH---EF 452
Db 582 DROMPKIEYRKIEVEDYSLPMQLKDPATFTDTHAYPLLLVVDGTPGQSQSSEFEVWTET 641
QY 453 QVLAAGYAVYINPRSGHYGOEFNAVVRGDDYGGKDDVMOAVDEAKRPHIDPKRL 512
Db 642 VLVSSHGAVVWCDGRSGFGQTKLLHVEVRRLLGFLEKDQMEAVTMLK-EQYIDKTRV 700
QY 513 GVTGSGYGFMTNIV-----GQT-----NRFAAQTORSISNWSIFHGV 552
Db 701 AVFGDYGGILSTYILPAKGENOGQFTTCGSALSPTIDFKLYASAFSE-YYLGLHGL 756
QY 553 SDIGYFTDQLEHDMFEDTEKLWDRSPLKYAANVETPLLIHGERDDRCPIEQAEQLFI 612
Db 757 DNRAEYMT--KLAH-----RVSAEDQOQLIITHATADEKIHFOHTAELIT 799
QY 613 ALKMKGKETKLVRFPNASH 631
Db 800 QLIKGNKANSYLSQIYPDESH 818

RESULT 11
ID DPP6_HUMAN STANDARD; PRT; 865 AA.
AC P42658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIPEPTIDYL PEPTIDASE IV LIKE PROTEIN (DIPEPTIDYL AMINOPEPTIDASE-
DE RELATED PROTEIN) (DIPEPTIDYL PEPTIDASE VI) (DPPX).
GN DPP6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=93372805; PubMed=8103397;
RA Yokotani N., Doi K., Wentholt R.J., Wada K.;
RT "Non-conservation of a catalytic residue in a dipeptidyl
RT aminopeptidase IV-related protein encoded by a gene on human
RL chromosome 7.";
RL Hum. Mol. Genet. 2:1037-1039(1993).
CC !- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC !- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC !- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPPX-L (SHOWN HERE) AND DPPX-S;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC !- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC PROLYL OLIGOPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M96859; AAA35760.1;

```

```

DR EMBL; M96860; AAA35761.1;
DR MEROPS; S09_973;
DR MIM; 126141;
DR InterPro; IPR001375;
DR InterPro; IPR002469;
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; peptidase_S9; 1.
KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 95
FT TRANSMEM 96 116
FT FT
FT FT
FT DOMAIN 117 865
FT CARBOHYD 173 173
FT CARBOHYD 319 319
FT CARBOHYD 404 404
FT CARBOHYD 471 471
FT CARBOHYD 535 535
FT CARBOHYD 566 566
FT CARBOHYD 813 813
FT VARSPLIC 1 81
FT FT
FT FT
FT FT
SQ SEQUENCE 865 AA; 97588 MW; 14BLAE0E0024464B CRC64;

Query Match 6.4%; Score 224; DB 1; Length 865;
Best Local Similarity 20.1%; Pred No. 4.7e-07;
Matches 165; Conservative 117; Mismatches 263; Indels 274; Gaps 40;

QY 2 KKLITADDI-TAIVSVTDPOYAPDQGTAAAYVK-SQVNEKDSYTSNIWIYETKTGSGVPW 59
Db 128 KKKVTVEDLFSDDFKIHDP-----AKWISDTFIYREOKGTVRLNVEINT--STVL 178
QY 60 THGEK-----RSTDPWSPDGRTLAFISDRECDAAQLYINSTEGGARKLTDIPYGVSKPL 115
Db 179 IEGKKIESLRALRYEISPDREYALFSYVVE-----PIYQHSYTYGYV--LSKIPHGDPQL 232
QY 116 -----WSPDGESILVTISLGESESIDREKTEODSYEPVEV--QGLSYKRD 159
Db 233 DPPEVSNAKLOYAGWGPQGQLIFIF-----ENNIIYCAHVGAQATRVST 278
QY 160 GK-----GLTRGAYAQVLVSVKSGEMKELTSHKADHGDFAFSPDGKWLVSANLTETD 213
Db 279 GKEGVINGLSLDWLYEEILAT-----HIAHWWSPDGTRLAYAA-----IN 319
QY 214 DASKPHDVIYIMSLE--SGDL-KQVTP-HRGSFGSSSPDGRYLLALIG----- 257
Db 320 DSRVP-----IMELPTYTGSIVPTVKPYHYPKAGSENPISLHVICLNGPTHDLEMMPPDD 375
QY 258 -NEKEY-----KNATLSK--AWLYDIEQGRLTCLTEMLDVHLADALIG----- 298
Db 376 PRREYIITWVKWATSKVAVTWNRAQ--NVSILT-----LCDATTGVCCTKKHEDESE 427
QY 299 SLIGGAQRPIWTKDSQGFY-----VPIRLEKEYIN-----SFLSLP 355
Db 428 AWHLRQNEPEVSKDGRKFFFIAPGGRGKFKYHIVSSQPNSSNDNIOSITSGDWDV 487
QY 319 --VIGTDOGSTGIYISIEGL-----VPIRLEKEYIN-----SFLSLP 355
Db 488 TKILAYDEKGNKYFLSTEDLPRRQLYSANTEGNFNRQCLSDLVENCITYFSASFISM 547
QY 356 DEQHFIASTVTKDRP-----SELYSIPLGOEKEQLTGANDKFKVREHTISIP-EI 404
Db 548 D--FLLKCEGCPVPMVTVNTDDKKRMFDLETNEHVKK--AINDR-----QMPKVEY 596
QY 405 QVATEDGVNMGVWMLMRPAQMEGETTYPLILNIHGPHMYGHTYFHEFOV-----LAAK 458
Db 597 RDEIDDDYNLPMQILKPAFTFTDTHYPLLLVVDGTPG---SOSVAEKFEVSWETVMSVSH 653
QY 459 GYAVVYINPRSGHYGOEFNAVVRGDDYGGKDDVMOAVDEAKRPHIDPKRLGVTGGS 518
Db 654 GAVVVKCDGRSGFGQTKLLHVEVRRLLGLLEEKDOMEAVRTMLK-EQYIDTRAVVFGKD 712

```

QY 519 YGGFWMTNIV-----GOT-----NRPKAAVTRORSINWISPHGVSDIGVF 558
|||:::|LPAKGENOGQTFCGSALSPTDFKLKYASAFSE---YLGHLGLONRATE 768
Db 713 YGYLSITILPAGENOGQTFCGSALSPTDFKLKYASAFSE---YLGHLGLONRATE 768

QY 559 FTDWQLHDMFDETEKLVDRSPPLKYAANVETPLLIIHGGERDDRCPIEQAEQIFIALKKMG 618
|:::|-----:|LHGERDDRCPIEQAEQIFIALKKMG 618
Db 769 MT--KVAAH-----RVSALFEQQFLIHPATADEKIHFQHTAELITQLIRCK 811
|:::|-----:|LHGERDDRCPIEQAEQIFIALKKMG 618

QY 619 KETKLVFRPNASHNLSTRGHPRQRIRKRLNIYSWFDOHL 657
|:::|-----YFTSSSLKQHL 835
Db 812 ANYSLIQYPDES-----YFTSSSLKQHL 835

RESULT 12
DAP2_YEAST STANDARD; PRT; 818 AA.
AC PI8962;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIPEPTIDYL AMINOPEPTIDASE B (EC 3.4.14.-) (DRAP B) (YSCV).
GN DAP2 OR YHR028C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[]
RN SEQUENCE FROM N.A.
RP MEDLINE=89174971; PubMed=2647766;
RX Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RT "Structure, biosynthesis, and localization of dipeptidyl
aminopeptidase B, an integral membrane glycoprotein of the yeast
vacuole";
RT J. Cell Biol. 108:1363-1373(1989).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
Vaughan D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
Vignati M.;
RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RT Science 265:2077-2082(1994).
RL -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYXOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC PROLYL OLIGOPEPTIDASE FAMILY. STRONG. TO DRAP A.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in part modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annotation/send_email_to_license@isb-sib.ch).
EMBL; X15484; CAJ33512.1; -;
DR EMBL; U10399; AAB68879.1; -;
DR PIR; A30107; A30107.
DR PIR; S46780; S46780.
DR MEROPS; S09.006; -;
DR SGD; S0001070; DAP2.
DR InterPro; IPR001375; -;
DR InterPro; IPR002469; -;
DR InterPro; IPR002471; -;
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1479;
 RX MEDLINE=97080550; PubMed=8921895;
 RA Sen K., Sikkema D.J., Murphy T.F.;
 RT "Isolation and characterization of the Haemophilus influenzae tolQ,
 RT tolR, tolA and tolB genes.";
 RL Gene 178:75-81(1996).
 CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE TOLB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32722; AAC22040.1; -;
 DR EMBL; U32470; AAC44597.1; -;
 DR TIGR; HI0382; -;
 KW Transport; Protein transport; Periplasmic; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 427
 FT VARIANT 6 6
 FT VARIANT 14 14
 FT VARIANT 17 19
 FT VARIANT 21 21
 FT VARIANT 79 79
 FT VARIANT 129 129
 FT VARIANT 160 160
 FT VARIANT 237 237
 FT VARIANT 322 322
 FT VARIANT 326 326
 FT VARIANT 328 328
 FT VARIANT 427 AA; 44967 MW; 0882201ABE9254B9 CRC64;
 SQ SEQUENCE
 Query Match 5.4%; Score 190; DB 1; Length 427;
 Best Local Similarity 24.4%; Pred. No. 2.8e-05;
 Matches 80; Conservative 32; Mismatches 102; Indels 114; Gaps 13;
 QY 26 TRAAVVKSVQVCKDYSNIIWYETKGGVPPWTHGKSRSTD----- 68
 DB 167 TRIATVWQ-----KNGSQPY---EVRVADYDGYNQFIVNRSQAQIM 205
 QY 69 -PRWSPDGRTLAFISDREGDAQAQYIMSTEGGAEARKLTDPYGVSKPLWSPDGESILVTI 127
 DB 206 SPWSPDGRQALVYS-FENKKSQLVVQDLNSGARKWASFOGHNGAPAFSPDGSRALAFAS 264
 QY 128 SLGEGESIDDRKTEQDSDYEPVEVOGLSKYKRDGKGLTRCAYAQVLVSVKSGEMKELTSH 187

Db 265 S-----RDG-----VLIYVMGANGGTPTLTSG 288
 QY 188 KADHGDPAFSPDGKWLVSANLTETDDASKPHDVYIMSLESGDLKQVTPHRGSGSSSFS 247
 Db 289 AGNTEPAWSPDGNLSILF-----TSDRSGSPQVYRMDASGGSATAV-----GGRGSAQIS 338
 QY 248 PDGRYIALL-GNEKEYK-----NATLSKAWL-----YDIBQGRLTCLITE 285
 Db 339 ADGKTLVIMNGNNVVKQDITGVSEVLSFGLGSPSPNGIMLIYSYTOG-----LGK 394
 QY 286 MLDVHLADALIGDSLIG--GAQRPIWT 311
 Db 395 VLQVLSADGRFKASLPGSDGVKPPAWS 422
 RESULT 15
 PPCE_PIG ID PPCE_PIG STANDARD; PRT; 710 AA.
 AC P23687;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (POST-PROLINE CLEAVING ENZYME)
 DE (PE).
 GN PREP.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ACTIVE SITE SER-554.
 RC TISSUE=Brain;
 RX MEDLINE=91152034; PubMed=1900195;
 RA Rennex D., Hemmings B.A., Hofsteenge J., Stone S.R.;
 RT "cDNA cloning of porcine brain prolyl endopeptidase and
 RT identification of the active-site seryl residue.";
 RL Biochemistry 30:2195-2203(1991).
 RN [2]
 RP ACTIVE SITE HIS-680.
 RX MEDLINE=91291146; PubMed=2064618;
 RA Stone S.R., Rennex D., Wikstrom P., Shaw E., Hofsteenge J.;
 RT "Inactivation of prolyl endopeptidase by a peptidylchloromethane.
 RT Kinetics of inactivation and identification of sites of
 RT modification.";
 RL Biochem. J. 276:837-840(1991).
 CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
 CC RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
 CC ACIDS LONG.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
 CC OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: IN ALL TISSUES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M64227; AAA31110.1; -;
 DR PIR; A37942; A37942.
 DR MEROPS; S09.001; -;
 DR InterPro; IPR001375; -;
 DR InterPro; IPR002470; -;
 DR InterPro; IPR002471; -;
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PRINTS; PR00862; PROLIGOPTASE.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 03:01:08 : Search time 3698.58 Seconds
(without alignments)
8242.875 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaagagctgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :
1: gb_ba1 : *
2: gb_ba2 : *
3: gb_ba3 : *
4: gb_in1 : *
5: gb_in2 : *
6: gb_in3 : *
7: gb_om : *
8: gb_ov : *
9: gb_pat1 : *
10: gb_pat2 : *
11: gb_ph : *
12: gb_p11 : *
13: gb_p12 : *
14: gb_p13 : *
15: gb_p14 : *
16: em_ba1 : *
17: em_ba2 : *
18: em_fun : *
19: em_htgo_hum : *
20: em_htgo_inv : *
21: em_htgo_rod : *
22: em_htg_hum1 : *
23: em_htg_hum2 : *
24: em_htg_hum3 : *
25: em_htg_hum4 : *
26: em_htg_hum5 : *
27: em_htg_hum6 : *
28: em_htg_hum7 : *
29: em_htg_hum8 : *
30: em_htg_inv1 : *
31: em_htg_inv2 : *
32: em_htg_other : *
33: em_htg_rod : *
34: em_hum1 : *
35: em_hum2 : *
36: em_hum3 : *
37: em_hum4 : *
38: em_hum5 : *
39: em_hum6 : *
40: em_hum7 : *
41: em_in : *
42: em_om : *
43: em_or : *

44: em_ov : *
45: em_pat : *
46: em_ph : *
47: em_p1 : *
48: em_ro : *
49: em_sts : *
50: em_sy : *
51: em_un : *
52: em_v1 : *
53: gb_sts1 : *
54: gb_sts2 : *
55: gb_sts3 : *
56: gb_sy : *
57: gb_un : *
58: gb_v11 : *
59: gb_v12 : *
60: gb_htg1 : *
61: gb_htg2 : *
62: gb_htg3 : *
63: gb_htg4 : *
64: gb_htg5 : *
65: gb_htg6 : *
66: gb_htg7 : *
67: gb_htg8 : *
68: gb_htg9 : *
69: gb_htg10 : *
70: gb_htg11 : *
71: gb_htg12 : *
72: gb_htg13 : *
73: gb_htg14 : *
74: gb_htg15 : *
75: gb_htg16 : *
76: gb_htg17 : *
77: gb_htg18 : *
78: gb_htg19 : *
79: gb_htg20 : *
80: gb_htg21 : *
81: gb_htg22 : *
82: gb_htg23 : *
83: gb_htg24 : *
84: gb_htg25 : *
85: gb_pr1 : *
86: gb_pr2 : *
87: gb_pr3 : *
88: gb_pr4 : *
89: gb_pr5 : *
90: gb_pr6 : *
91: gb_pr7 : *
92: gb_pr8 : *
93: gb_pr9 : *
94: gb_rol : *
95: gb_rod : *
96: gb_in4 : *
97: gb_pr10 : *
98: em_ba3 : *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match %	Length	DB	ID	Description
1	1971	100.0	1971	9	AX006271	AX006271 Sequence
C	1971	100.0	217420	2	BSUB0017	299120 Bacillus su
C	238.2	12.1	265118	2	CNSPAX06	AJ248288 Pyrococcus
C	238.2	12.1	265118	9	AX041922	AX041922 Sequence
5	227.2	11.5	1896	10	EL6634	E16634 Pyrococcus
6	227.2	11.5	233000	2	AP000003	AP000003 Pyrococcus
C	191.2	9.7	10256	1	AE001879	AE001879 Deinococc
8	161.8	8.2	251700	2	AP000062	AP000062 Aeropyrum

9	158	8.0	304290	2	CNSPAX05	AJ248287	Pyrococcus
10	158	8.0	349980	9	AX041921	AX041921	Sequence
11	150.6	7.6	257000	2	AP000002	AP000002	Pyrococcus
12	149.8	7.6	1869	9	AR086952	AR086952	Sequence
13	149.8	7.6	1869	9	AR095647	AR095647	Sequence
14	132.6	6.7	343800	2	AP000993	AP000993	Thermopla
15	130.4	6.6	9825	1	AE005114	AE005114	Halobacte
16	130.4	6.6	247695	2	AP000064	AP000064	Aeropyrum
17	75.2	3.8	139203	73	AC068311	AC068311	Homo sapi
18	65	3.3	1998	5	AF059278	AF059278	Entamoeba
19	64.8	3.3	2400	2	BSTHBC	AF04603	B. subtilis
20	62.8	3.2	1992	5	AF059279	AF059279	Entamoeba
21	54.4	2.8	7218	10	I66494	I66494	Sequence 14
22	46.4	2.4	67965	3	SOLRPN09	AL512970	Sulfolobu
23	46.2	2.3	316900	3	TACID3	AL445065	Thermopla
24	46	2.3	2325	12	AF082514	AF082514	Trichophy
25	45.4	2.3	13928	1	AE004038	AE004038	Xylella f
26	41.6	2.1	3406	2	AF299085	AF299085	Myxococcu
27	41.2	2.1	38000	15	SPBIC1711	E32987	Gene encodi
28	39	2.0	6741	10	E32987	AE006030	Caulobact
29	39	2.0	11452	1	AE006030	AC073864	Homo sapi
30	39	2.0	136257	88	AC073864	AC026678	Homo sapi
31	39	2.0	170075	70	AP006678	AF000852	Homo sapi
32	39	2.0	205400	82	AP000852	AC024935	Homo sapi
33	39	2.0	313462	69	AC024935	AE005872	Caulobact
34	38.4	1.9	13212	1	AE005872	AC090315	Homo sapi
35	38.4	1.9	57223	77	AC090315	AC025602	Homo sapi
36	38.4	1.9	175035	69	AC025602	U100044	Drosophil
37	38	1.9	119844	61	AC010044	U13070	Caenorhabdi
38	37.6	1.9	40824	6	CELF01F1	AC021568	Homo sapi
39	37.6	1.9	162454	66	AC021568	AC087878	Homo sapi
40	37.6	1.9	166393	7	AC087878	AC011710	Homo sapi
41	37.6	1.9	191637	62	AC011710	AC018475	Homo sapi
42	37.6	1.9	204879	65	AC018475	E02747	DNA encodin
43	37.4	1.9	2196	10	E02747	D00524	Sus scrofa
44	37.4	1.9	2385	7	PICAAAREL	AC084049	Homo sapi
45	37.4	1.9	169950	76	AC084049		

RESULT	1	1971 bp	DNA	PAT	24-AUG-2000
AX006271	Sequence 1 from Patent WO9903984.				
CCUS	DEFINITION				
AX006271	ACCESSION				
AX006271.1	VERSION	GI:9929124			
KEYWORDS					
AX006271	SOURCE				
AX006271	ORGANISM				
AX006271		Bacillus subtilis.			
AX006271		Bacillus subtilis			
AX006271		Bacteria; Firmicutes; Bacillus/Clostridium group;			
AX006271		Bacillus/Staphylococcus group; Bacillus.			
AX006271		1 (bases 1 to 1971)			
AX006271		Estell, D.A.			
AX006271		Proteases from gram-positive organisms			
AX006271		Patent: WO 9903984-A 1 28-JAN-1999;			
AX006271		GENECOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)			
AX006271		Location/Qualifiers			
AX006271		1..1971			
AX006271		/organism="Bacillus subtilis"			
AX006271		/db_xref="taxon:1423"			
AX006271		558 a 431 c 524 g 458 t			
AX006271		source			
AX006271		BASE COUNT			
AX006271		ORIGIN			
AX006271		Query Match	100.0%;	Score 1971; DB 9; Length 1971;	
AX006271		Best Local Similarity	100.0%;	Pred. No. 0;	
AX006271		Matches 1971; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;	
AX006271		1 atgaaaagctgataaccgcagacacacacagcagattgtctgtgaccgatcctcaa	60		
AX006271					
AX006271		1 atgaaaagctgataaccgcagacacacacagcagattgtctgtgaccgatcctcaa	60		
AX006271					

QY 1141 gaagagaacacagctgactgagcgcgaatgacaagtgttgcaggagacatacgaata 1200
 DB |||||||
 DB 1141 GAAGAGAAACAGCTGACTGCGCGAATGACAAAGTTTCTCAGGAGGATACGATATCAATA 1200
 QY 1201 cctgaagagattcaataatctactacagaagcggcgtgatggtgaacgctggctgagag 1260
 DB |||||||
 DB 1201 CTTGAAGAGATTCAATATCTCTACAGAGACCGCGCTGATGTCGAACGGCTGGCTGATGAGG 1260
 QY 1261 cctgcacaaatggaagtgtagacaacataccattcttcttaacatacacacgagctccg 1320
 DB |||||||
 DB 1261 CTTGCACAAATGAAGTGCAGACACATATCTTATTTCTTAACATACACGCGCTCGG 1320
 QY 1321 catatgatgcagacatacatatttcatgagtttcaggtctggtgagcggcgaagagatac 1380
 DB |||||||
 DB 1321 CATATGATGTCAGGACATACATATTTTCATGATTTCAGGTCTGGCGCGGAAGGATAC 1380
 QY 1381 gggctcgctttatcaatccagagagaagccacgctacggcagcaggaatttggatgagct 1440
 DB |||||||
 DB 1381 GCGGTCTGTTATATCAATCCGAGAGAGACCGCTACGGCTACGGCAGGAAATTTGTGAATGCG 1440
 QY 1441 gtcagagagattatgggggaaagattatgacgatgtgacgagcgtgagcagcgtgagcagct 1500
 DB |||||||
 DB 1441 GTCAGAGAGATTATGGGGGAAAGGATTATGACGATGTCATCGAGCTGTGGATGAGGCT 1500
 QY 1501 atcaaacagatcccatattgatctactaaagcggctcggtgacagcggcgaagctacgga 1560
 DB |||||||
 DB 1501 ATCAAACAGATCCCATATTGATCTTAAGCGGCTCGGTGTCACGGCGGAAGCTACCGGA 1560
 QY 1561 ggtttatgacaaactgagctgctgagcagcagcagcagcagcagcagcagcagcagcagc 1620
 DB |||||||
 DB 1561 GGTTTTATGACCAACTGATGCTCGGCGAGACGAAACCGCTTTAAGCTGCGCTTACCCAG 1620
 QY 1621 cgctcgatatcaaatggatgcagcttcaagcggctcagtgatgcagctgagctattcttaca 1680
 DB |||||||
 DB 1621 CGCTCGATATCAAAATGGATGAGCTTTTACCGGCTCAGTCATATCGCTATTTCTTTACA 1680
 QY 1681 gactggcagcttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
 DB |||||||
 DB 1681 GACTGGCAGCTTGAGCATGATGCTTTTGAGGACACACAAAGCTCTGGGACCGGTCTCTCT 1740
 QY 1741 ttaaaatacagcagcaaacgctggagacaccgctttgtactgacgagcagcagcagcagcagc 1800
 DB |||||||
 DB 1741 TTAATAATACGACCAACAGTGGAGACACCGCTTTTGATGATGCTGATGCGAGCGGATGAC 1800
 QY 1801 cgatgccgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1860
 DB |||||||
 DB 1801 CGATGCCGATGAGCAGCGGAGCAGCTGTTATCGCTCTGAAAAAATGGGCAAGGAA 1860
 QY 1861 accaagcttgcgctttccggaatgcatgcgacaaatttatcacgacccgacaccccaaga 1920
 DB |||||||
 DB 1861 ACCAAGCTTGTCCGTTTCCGAATGATGCGACAAATTTATCAGCGACCGGACACCCCAAGA 1920
 QY 1921 cagcgatcaaacgctgatttatcagctcagctcagctcagctcagctcagctcagctcagctc 1971
 DB |||||||
 DB 1921 CAGCGATCAAGCGCTGAATTATATATCAGCTCATGCTGTTTGTATCAACATCTC 1971

RESULT 2

BSUB0017/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BACILLUS subtilis complete genome (section 17 of 21): from 3197001 to 3414420.
 299120 AL009126
 299120.1 GI:2635613
 Bacillus subtilis.
 Bacillus subtilis.
 Bacteria: Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus.
 1 (bases 1 to 217420)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriess, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,

Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerston, P.T., Entian, K.D., Errington, J., Fabre, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligorsky, A., Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Portet, D., Porwoll, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Viaria, A., Wambutt, R., Wiedler, E., Wiedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

Nature 390 (6657), 249-256 (1997)

2 (bases 1 to 217420)

Direct Submission

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,

Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES

source

Location/Qualifiers

1. 217420
 /organism="Bacillus subtilis"
 /strain="168"
 /db_xref="taxon:1423"
 complement(3..1277)
 /gene="yule"
 complement(3..1277)
 /gene="yule"
 /function="unknown"
 /note="similar to L-rhamnose isomerase"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15107.1"
 /db_xref="GI:2635614"
 /translation="MTIKANYDSAKOAEKVGIDVEEALROLEQVDPISLHCWGDDIE
 GFEVNGKLSGGIDVTGNYPGKAQTPPELRDREKXLSLIPGKRVNHLAIYAETNRE
 AVERDELKPOHFENWVKAQNLGLDFNFTLSHEKAADGLTSHHPDPIRFEFTWRH
 CIACRICEYFGKELGTPCLTNIPWIDYKIDPSDLTPRKLKESLDRIFFSEITSEQ
 HNLDSIEKLFSGSESVVVGSHFYLALYLNHKLCLDLDGTGHFHTPTVSNKISSML
 LYTDIALHVSPPWSDHVVVLDDELREIALEIVRNHAKLEKVAIGLDFDFAINRV
 AAWTIGTRMKALYLLIPNGYLKQLEEGRYTERLALMEEFKTYPPGAIWDSYCE
 QMGVPKAWLYDIKEYEQVLLKRAKSSPIV"
 complement(1303..1617)
 /gene="yuld"
 complement(1303..1617)
 /gene="yuld"
 /function="unknown"
 /note="similar to hypothetical proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB15108.1"
 /db_xref="GI:2635615"

<p>CDS</p> <p>YVGPFGILNPKAAKSVIMHQTIAHARAYDSIKFTSDKPVGVILNLIPIAYPLNPRDSK HYKAAENYDLFHNRLFLEAVNGKGLDIDINGEVYKVPKLRNDWIGNNYITREVVKYV EPKYELPLVTVGVGVYGYSANPNSISDNPTSDFGWEVYKGVYKGLYDSTABAYEYSE MYFITENGADSKDILRLPYIYVDHVKVRKLRRENGVNLGYFHWALTDNYEWAMGFYK RGLYEVDPITIKERIPRRKSVETYKKVKEGIE"</p> <p>1834. .2283</p> <p>/note="PAB1019"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50448.1"</p> <p>/db_xref="GI:5458962"</p> <p>/translation="MPLPPDITFDSVTLLKMHSEKRRLEITLAKFTVLSMTVQR YLPARAYLKNVELEFVKLDIYNIPVSEIILRSKIANLIRKGVFLDLEIITA VTAETGSLITDEPKKYEPYIRRGGLDTMPDLKFLREVELMVEKEII"</p> <p>complement(2276. .3079)</p> <p>/note="PAB1341"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50449.1"</p> <p>/db_xref="GI:5458963"</p> <p>/translation="MPENLESFEEAFKLLIEVLEFDLNDPDKVVKVLCIEPHDDC VIGGGTICKLTERGEIVITYACMTDGTGNGTLDSSLTGHELATIRREDEESSKLLGVK KIYLNTRDTELPYSREVRKDLVRIIRKRPDGVFLPDPLWLPYEAHPDRHNTGLALD AVAFSPLPNFNSVDVEIGLGHQVFSIYALYTNPNFVDLIIDVWELKLAIRTHQS FPDQWVEWEPFLRTVALFYKAGVYAEGRIMPGLFYHITPFADLI"</p> <p>3192. 4340</p> <p>/note="PAB1020"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50450.1"</p> <p>/db_xref="GI:5458964"</p> <p>/translation="MKEVSVSSVYKEILVKLGITEDIETLEMKGGIIEDEFDGIRYVR PKDSAGLRGVTVDEEVYIPGFPHIKRIINLRGIRIFRKRGEYVEEKVDGVNVR VYVYKGMIGITRGFGICPFTTPIPDVFPQPFQFNDPNLILVGMAGPESYLVGEP PIVKEDIOFFLDQVEIKTGRSLPVEERLKAIEYGINHVEYFGYTKDDVDLYOLI EHLSEGRGIIIMKSPDMKKIVKYVTPYANINDIKIGARVYFVLPYGVTSRISLAF YLAERKIRGEFEERVAKELGSLALLOPFVESIDVEQEEIDHELFKVRVKRIETAYKMV THPEKLGKIEIVDIEEIKDGNRIITFRKRIYDPATNEIRLEIGGKAFVD"</p> <p>4351. 5658</p> <p>/note="PAB1021"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50451.1"</p> <p>/db_xref="GI:5458965"</p> <p>/translation="MEDIALRLVKAIRAADPYKAVINSVRVSGNLLVKDKKEFEIKG KVVLAFGKAACDMARAIEIILDEVDGAVTKYGYKELAKRIEAGHPIDENSIR GATEGLKLLKVEKEEDIVFLISGGSAFPELDEGLIEDLKLTLNELLKSGAKIHE INTYRKHISVKVGGKLAKEIGTGIVLIIISDVVDNLEIASGTVKDPDTFODAKRI LELYGWDRPVSRSYHRIKIGEAETLKEELPNHNLFIASNSISQAVVEEATK LGPKYVMTTLTGEAKDAGLFGISVIOETAKTGNPFPPVLPVFGGTTVTIEGEG KGGPNQELASATRKLNLAIVAFDITDGTGTPDAAGGIVDGETVEKLRKRGIDIE KVLKHNYSALKKVGLSLFTGPTGNVNSMIILITSKGRI"</p> <p>complement(5633. .6655)</p> <p>/note="PAB1338"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50452.1"</p> <p>/db_xref="GI:5458966"</p> <p>/translation="MPIYERARKVGVIGSSWVTSKVSREDLAOLESDRIYVVKLRMGE VTGDSFPFSALTDIGIRIPDELKLNPEKDFHGLKFFYVFGGELLVIFTIEENLIYD BKKAEPFVKFVFLVINGKTVMLQARIVGGAANNESKWEGLVRKIVKASRTQRIE RSVVITQDKRPVSIFDLEIEIEFVEMNGKVRANKRIHFHINQSVISLYIPDKK VOLFLVRLLYLTPSAMFEIILKADDPPLTKSEFOELMEKLELESDEMEKQLLLA LYSGINDELHQFLGITEKEIEEIVDRMIDKGLLKIVMIRKIVDLTNGRKIVANKLK YGLVSM"</p> <p>complement(6670. .7425)</p> <p>/note="PAB1337"</p> <p>/codon_start=1</p>	<p>CDS</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="CAB50453.1"</p> <p>/db_xref="GI:5458967"</p> <p>/translation="MDVGHILITAVVVLIGSVISAMTVSSVISOQMKMCAALMISRSL EVETKTKRRKTRGSEAKPEVAKQEVDRVSEEDIVRKIOLIDEKVINVAEAKK KRLRLMLLDVARGYALGFVSEDEYNFALMRVISELDEFKRLWLAAYPFPEDRLNQ MLTHVARTLPFIEVREKKGEGSVKLSPLDIALIRITTSIKSAISILDDLIKRRGDNP AVTPLEIKSQELEKCKKKVEDLEKRLQLEAL"</p> <p>complement(7468. .9687)</p> <p>/note="PAB1336"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="methyl-accepting chemotaxis protein (tlpC)"</p> <p>/protein_id="CAB50454.1"</p> <p>/db_xref="GI:5458968"</p> <p>/translation="MOPFRKKIILMAINSAVITMAIFGVTIYSFNKVVYIKGEIQAP AIEQGGLAMSAEVAEKMPDDFARVADYGTMAEYVVKAYAKGLDSNETAMREFLL SEFKIKELNKNDYSIVTFGDEGHMYWPPDELPEGVDPVRPMYVOAKKGGALITE PYRDAATQWVITYAVPVNGQFKVIGLDVFSVTLNETKKIKLGTGYIAIIDQN GMTLVHPNEEFINKLNIQVPELKDADLAKKDEGHVLTFFQIKKIGYKRLKTT GIVVATVPLSDITGPIILSATDTAINKMKMVIQGFVGVFVIVAILVYIKIASNTL APLEKLYAQAALAEGRKQVSEYKQIRYLERDEIGALIOAFEAESKDLVGTLNALIS KLERLAGELSLNGLTVEVREGELDIIEDIRSVETTFESIGSVIEMANDLEKRNAL AQKSDVTEANQVNEALQOVSIEAQROQETINETIDGMRLVAQTSESVRAMEEFSG AVTEVVSITANGSOGKDBALAKIEDIOHMSRIEFTVSKVAEMSRIEITVITSIA EQTNLLALNAIEAARAGEAGRFVAAQERKLAESKQADNLIKSIIDKITDEIKE AVEATKEGVSIGESSETLRTDTIGYLANIATILLQETSERMTVKEQIVRTQEVKAL RALENLAASABETTASABEVSASIEQQTAAIEELRAAQELKDMYGRMRQIVGKFKV"</p> <p>complement(9704. .10195)</p> <p>/note="PAB1335"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="CHEMOTAXIS PROTEIN CHED"</p> <p>/protein_id="CAB50455.1"</p> <p>/db_xref="GI:5458969"</p> <p>/translation="MTREIKVIGYAVYKGEIISTYGLGSCVGLTYDRVTYKVGGL LHALLPEAARYGHRGNPAKYVDVTGLQLLKEVLKLGASKRFLKALFUGAQMFPQNIK DELKIGERNVQAKRELKGLRLVAEDTGGRGRTYLDLSTGKVRMRKVIQGVQIE KYV"</p> <p>complement(10179. .10817)</p> <p>/note="PAB1334"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="CHEA HISTIDINE KINASE"</p> <p>/protein_id="CAB50456.1"</p> <p>/db_xref="GI:5458970"</p> <p>/translation="MKKSLEYKDFKEASNIAMSHALTSLSQIMIGGPIEMEAPDVEIV SRAEFLKLLAERGYSKGFVMDVTGSLGTILOPKHSAIINISAYIMGMEGSEEE LDEMGSAMEVGNILISATDILSNLIGEPVSLSPKPAESLYDIKEKLGRLPDRNV NSIVPFSKSEYKEDIGVESFYLYVPTPESTFLVKLEKQISEVEANDEGN"</p> <p>complement(10814. .11452)</p> <p>/note="PAB1333"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="CHEA HISTIDINE KINASE related protein"</p> <p>/protein_id="CAB50457.1"</p> <p>/db_xref="GI:5458971"</p> <p>/translation="MDIVAENIEEYVKNLDFEAKSALLETNIGASHAATAALSEMTGR EVNISPDLKVVKEIKNVEIPEVGVVYIELGKDFSSHAFNACFADDDVLRMDVIM GNPPTTTELDMDVKSSPMENILISAFANALSOFLGITEQPPALTIDFLPALID LALADJGKCDYTILOTDIKISGVFEFHFLLFPKPEDMKKILEKLLGLLA"</p> <p>complement(11436. .13754)</p> <p>/note="PAB1332"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="chemotaxis histidine kinase (cheA)"</p> <p>/protein_id="CAB50458.1"</p> <p>/db_xref="GI:5458972"</p> <p>/translation="NVTLEDMSOYLDEFLADARDRIDSNAILTEKIVKEGSEEE KAEILNQIFRQAHTKTAATMGKNLSETAHKMENLFDARNSQIETLTPPEVDVILD FLDAEAMVNNEIESGSGDDIVSPLFRADFEILOGGKGAEKKPEEKPEEVAKEE EKPEKVEEGVEGKVYHIKVIYFOKDAPLKVRAFLILSDLEERGEVITWNPGRDVISSG</p>
--	---

	BASE COUNT	ORIGIN
/db_xref="taxon:32644"	614 a	358 c 491 g 433 t

Query Match 11.5%; Score 227.2; DB 10; Length 1896;
Best Local Similarity 55.4%; Pred. No. 7.5e-60;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;

QY	1139	aggaaagaaacacgctgactggcgcgaatgacaagtgtgtcaggagagacatcatca	1198
Db	1061	ATGGAAGAGGAGAAAGGTTACCGACTTTACAAATGGATAAAGGTTACACCCTTTCAA	1120
QY	1199	tacctgaagagattcaatgatctcacagaagcggcgtgatggtgaacggctgctgatga	1258
Db	1121	AACCTGAACACTTTAAGGTTAAAGCAAGTGACGGGTTGAAATAGATGCTCGGTAATGA	1180
QY	1259	ggcctgcacaaatggaaagtgagacaaacatatccacttattcttaacatacacgcgcgtc	1318
Db	1181	AACCGGTGAACCTTCAGSAAAGAAAGAGTAGTCCAGCTATTCTAGAGATCCACGGTGGTC	1240
QY	1319	cgcataatgatgcagagacatacatatttccatgagtttcagtgctggcggcgaagaagat	1378
Db	1241	CTAAAAACCGCTTACGGTTACGCTTTATGACAGAGTTCACGTTTAAAGCTCTAAAGGCT	1300
QY	1379	acgcggtgcttatcatcctcagagaagccacgcgtctacggcaggaatttgtgaatg	1438
Db	1301	TCGTCGTGATATTCTCAAAATCCTAGAGGAGGCGATGGCTACGGAGAGGAGTTCCCGAT-	1359
QY	1439	cggtcagagagattatgggaaagattatgaacgatgtgatcaggctgtgtgatgaggt	1498
Db	1360	--ATRAGGGACACTATGGGAGAGGGATTACAGAGATTAAATGGAGGTAGTCGATGAAG	1417
QY	1499	ctatacaacagatccgcataattgattcctaagcggctcgggtgtcaocgggcggaagctacg	1558
Db	1418	CATTAGGAGATTGACTTCATAGATGGGAAAGGCTAGGAGTTACCGGGGTTCCTATG	1477
QY	1559	gagggtttatgacaaactggatcgtcgggcagacgacacgcgttaaaagctgcggtaccc	1618
Db	1478	GTGGCTCATAGAGAACTGGATATAGTCGGACATACCAACAGGTTCAAAGCCGCTGTACCC	1537
QY	1619	agcgtcgatatacaatttgatcagctttcacggcgtcagtgatcogctattcttta	1678
Db	1538	AGAGATCAATTTCAAAITGGATAGCTTCTTCGGGACACCGGATATAGGTTATTACTTTG	1597
QY	1679	cagactggcgcgttgacatgacatgtttgaggacacagaaaacgtctgggacgcgctc	1738
Db	1598	CTCCAGATCAAAATAGSAAAGATCCCTGGAGCAACTTGGAAAGGTTATTTGGGAAAAGACC	1657
QY	1739	ctttaaataacgacgaacagtgagacacgcgttttgatactgcattgcgcgagcggatg	1798
Db	1658	CATTAAAGTAGCTCCCAACGTTGAAACTCCCCGTGCTTTAATCCACTCTACCGAAGACT	1717
QY	1799	accgatcccgatcgcagcagcgagcagctgtttatcgctctgaaaaaatgggcaaggt	1859
Db	1718	ACAGGTGTGGCTTCCCCAGGCAATTGCAACTTCTCATATCCCTAAATACCTTGGGGAAGA	1777
QY	1859	aaacaaagcttgctgttttccgaatgcacgcacaaatttatcgcgcgcgcgcacccaa	1917
Db	1778	GAGTTGAATTGGCAATATTCACAGAGAGAAATCATAGACTTAAGTAGCTGGGAAGCCAA	1837
QY	1919	gacgcgcgatcaagcgcgtgaattatatcagctcatgtgttgatcaaatct	1970

RESULT	6
AP000003	
LOCUS	
DEFINITION	Pyrococcus horikoshii OT3 genomic DNA, 544001-777000 nt. position (3/7).
ACCESSION	AP000003 AB009484 AB009485 AB009486 AB009487 AB009488 AB009489
	AB009490 AB009491 AB009492 AB009493 AB009494 AB009495
	AB009496 AB009497 AB009498 AB009499 AB009500 AB009501
	AB009502 AB009503 AB009504 AB009505 AB009506 AB009507
	AB009508 AB009509 AB009510 AB009511 AB009512 AB009513
	AB009514 AB009515 AB009516 AB009517 AB009518 AB009519
	AB009520 AB009521 AB009522 AB009523 AB009524 AB009525
	AB009526 AB009527 AB009528 AB009529 AB009530 AB009531
	AB009532 AB009533 AB009534 AB009535 AB009536 AB009537
	AB009538 AB009539 AB009540 AB009541 AB009542 AB009543
	AB009544 AB009545 AB009546 AB009547 AB009548 AB009549
	AB009550 AB009551 AB009552 AB009553 AB009554 AB009555
	AB009556 AB009557 AB009558 AB009559 AB009560 AB009561
	AB009562 AB009563 AB009564 AB009565 AB009566 AB009567
	AB009568 AB009569 AB009570 AB009571 AB009572 AB009573
	AB009574 AB009575 AB009576 AB009577 AB009578 AB009579
	AB009580 AB009581 AB009582 AB009583 AB009584 AB009585
	AB009586 AB009587 AB009588 AB009589 AB009590 AB009591
	AB009592 AB009593 AB009594 AB009595 AB009596 AB009597
	AB009598 AB009599 AB009600 AB009601 AB009602 AB009603
	AB009604 AB009605 AB009606 AB009607 AB009608 AB009609
	AB009610 AB009611 AB009612 AB009613 AB009614 AB009615
	AB009616 AB009617 AB009618 AB009619 AB009620 AB009621
	AB009622 AB009623 AB009624 AB009625 AB009626 AB009627
	AB009628 AB009629 AB009630 AB009631 AB009632 AB009633
	AB009634 AB009635 AB009636 AB009637 AB009638 AB009639
	AB009640 AB009641 AB009642 AB009643 AB009644 AB009645
	AB009646 AB009647 AB009648 AB009649 AB009650 AB009651
	AB009652 AB009653 AB009654 AB009655 AB009656 AB009657
	AB009658 AB009659 AB009660 AB009661 AB009662 AB009663
	AB009664 AB009665 AB009666 AB009667 AB009668 AB009669
	AB009670 AB009671 AB009672 AB009673 AB009674 AB009675
	AB009676 AB009677 AB009678 AB009679 AB009680 AB009681
	AB009682 AB009683 AB009684 AB009685 AB009686 AB009687
	AB009688 AB009689 AB009690 AB009691 AB009692 AB009693
	AB009694 AB009695 AB009696 AB009697 AB009698 AB009699
	AB009700 AB009701 AB009702 AB009703 AB009704 AB009705
	AB009706 AB009707 AB009708 AB009709 AB009710 AB009711
	AB009712 AB009713 AB009714 AB009715 AB009716 AB009717
	AB009718 AB009719 AB009720 AB009721 AB009722 AB009723
	AB009724 AB009725 AB009726 AB009727 AB009728 AB009729
	AB009730 AB009731 AB009732 AB009733 AB009734 AB009735
	AB009736 AB009737 AB009738 AB009739 AB009740 AB009741
	AB009742 AB009743 AB009744 AB009745 AB009746 AB009747
	AB009748 AB009749 AB009750 AB009751 AB009752 AB009753
	AB009754 AB009755 AB009756 AB009757 AB009758 AB009759
	AB009760 AB009761 AB009762 AB009763 AB009764 AB009765
	AB009766 AB009767 AB009768 AB009769 AB009770 AB009771
	AB009772 AB009773 AB009774 AB009775 AB009776 AB009777
	AB009778 AB009779 AB009780 AB009781 AB009782 AB009783
	AB009784 AB009785 AB009786 AB009787 AB009788 AB009789
	AB009790 AB009791 AB009792 AB009793 AB009794 AB009795
	AB009796 AB009797 AB009798 AB009799 AB009800 AB009801
	AB009802 AB009803 AB009804 AB009805 AB009806 AB009807
	AB009808 AB009809 AB009810 AB009811 AB009812 AB009813
	AB009814 AB009815 AB009816 AB009817 AB009818 AB009819
	AB009820 AB009821 AB009822 AB009823 AB009824 AB009825
	AB009826 AB009827 AB009828 AB009829 AB009830 AB009831
	AB009832 AB009833 AB009834 AB009835 AB009836 AB009837
	AB009838 AB009839 AB009840 AB009841 AB009842 AB009843
	AB009844 AB009845 AB009846 AB009847 AB009848 AB009849
	AB009850 AB009851 AB009852 AB009853 AB009854 AB009855
	AB009856 AB009857 AB009858 AB009859 AB009860 AB009861
	AB009862 AB009863 AB009864 AB009865 AB009866 AB009867
	AB009868 AB009869 AB009870 AB009871 AB009872 AB009873
	AB009874 AB009875 AB009876 AB009877 AB009878 AB00

[illegible]

RESULT	5
EI6634	1996 bp DNA PAT 28-JUL-1999
Locus	Pyrococcus horikoshi gene for aminoacylase/carboxypeptidase.
DEFINITION	
ACCESSION	EI6634
VERSION	EI6634.1 GI:57111317
KEYWORDS	JP 1998210977-A/1.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 1896) Ishikawa,K., Matsui,I., Ishida,H., Kosugi,Y. and Higuchi,K. THERMO-STABLE, ACYLPEPTIDE HYDROLASE, AND GENE CODING THE SAME Patent: JP 1998210977-A 11-AUG-1998; AGENCY OF IND SCIENCE & TECHNOL
COMMENT	OS Pyrococcus horikoshi PN JP 1998210977-A/1 PD 11-AUG-1998 PP 31-JAN-1997 JP 1997018381 PI ISHIKAWA KAZUHIKO, MATSUI IKUO, ISHIDA HIROYASU, PI KOSUGI YOSHIMI, PI HIGUCHI KATSUHIKO PC C12N15/09,C07H21/04,C12N9/78,C12Q1/37,(C12N15/09,C12R1:01), PC (C12N9/78, PC C12R1:19); CC strandedness: Double; CC topology: Linear; FH key Location/Qualifiers FT source 1..1896 ET /strain='JCM9574', FT mat_peptide 1..1896 ET /product='aminoacylase/carboxypeptidase'. Location/Qualifiers 1..1896 /organism='unidentified'
FEATURES	source

VERSION
KEYWORDSSOURCE
ORGANISMREFERENCE
AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE
AUTHORS

TITLE

JOURNAL

COMMENT

AP000003.1 GI:3236130

Pyrococcus horikoshii (strain:OT3) DNA.
Pyrococcus horikoshiiArchaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.

1 (sites)

Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kishida, N., Ouchi, A., Aoki, K., Nakamura, Y., Robb, T. F., Horikoshi, K., Masuchi, Y., Shizuya, H. and Kikuchi, H.
Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3
DNA Res. 5 (2), 55-76 (1998)
983441372 (bases 1 to 233000)
Tanaka, T., Kawarabayashi, Y. and Kikuchi, H.
Direct Submission

Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka

Kawarabayashi, National Institute of Technology and Evaluation,
Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo
151-0066, Japan (E-mail: genomeOT3@nitech.go.jp, Tel: +81-3-3481-8951,
Fax: +81-3-3481-8424)
On Jul 28, 1998 this sequence version replaced gi:3130861
gi:3130869 gi:3130912 gi:3130938 gi:3130941 gi:3130952 gi:3130999
gi:3131044 gi:3131081 gi:3131085.
AB009484-AB009494: submitted (10-DEC-1997)Kawarabayashi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.
Shizuya, H. is at the California Institute of Technology, Pasadena,
CA, USA.The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 103; EMBL release 52.0; SwissProt
release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genomeOT3@nitech.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: <http://www.bio.nite.go.jp/>.
Location/Qualifiers

1. 233000

/organism="Pyrococcus horikoshii"

/strain="OT3"

/db_xref="taxon:53953"

/map="544001-777000 nt. position"

complement(645..1067)

/gene="PH0605"

complement(645..1067)

/gene="PH0605"

/note="similar to PIR:A64312 percent identity: 51.111 in
136aa; PIR:A64312 percent identity: 51.111 in 136aa;
Swiss_Prot:P41375 percent identity: 44.762 in 108aa"

/codon_start=1

/transl_table=11

/product="140aa long hypothetical translation initiation
factor eIF-2 beta"

/protein_id="BAA29694.1"

/db_xref="GI:3257011"

/translation="MEIDYDYDEKLLLEKAEVPELVKHKSRFEVPGALVTIEGKNT
IIEFNKDIAELNRDQHLKLLRLREIATAGTLEGKRVVLQGRFTPLIANKLKKYIK
EYVICPVGSPDTKIKRORFYLKCEACGAETPIQHL"

gene

CDS

complement(1110..4436)

/gene="PH0606"

complement(1110..4436)

/gene="PH0606"

/note="similar to Swiss_Prot:P30665 percent
identity: 41.538 in 331aa. motif=APP/GTP-binding site motif
A (P-loop) ; protein splicing signature"

/codon_start=1

/product="1108aa long hypothetical cell division control
protein"

/protein_id="BAA29695.1"

/db_xref="GI:3257012"

/translation="MERREILIERFLFFREYAEAGEEPLYMGKIKDILLTTPKRSTVI
NMWILNSYDELAEEVIEPEECILAAEDAIQILKEDEFLREDVPVKIHFHNLPKTL
MYKEGAEHKNLIQVEGVVTRTEIKFVSVAVFVKDCGHEMVVPKQPYEGFVAVK
KCEQCGSKNVQDVEKSKFVNFQMRIDQRPESLRGGQMPRTDGLLLDDIVDTAMPG
DRVIVVGLIRVIOEKREKVPFKKIVFNHIEPLSKLELEISPEDEOKIRELAKRK
DIVSAIVDSIAPAIYGYKEVKKGIALALFGGVSRTLPGSTRLGDIHVLVLDGPGVAK
CVDYDEVILGGRKKGKIGEIVEEAKKAEKGRVDDGYAPINLEYALDVRTL
KYRVKADIAWKRTPEKMRIRIKRREIRVTPHPTFTEEGRIKTKKAYELKVG
KIATPREAPEAEIIFWDEVVEIEEYKPNNSVYDLQVPEHNFIAFGFVHNSOLLRY
VANLAPRAIYTGKSSAAGLAVPTLINTONGRVEICKFVEENKVEGEISEEGIS
YAPCRKVTFFKDGITVESPIRVKMLRAPKVLRIKTEGFSIALRETLLKLTNDG
ELSWAEAGVKVGYTVYKSEKVLPGAGKTIIDVSKLYNMMEVKDYLTRVEVKA
IEKLEIMNPMNKIPGQESYEELLKRLKETINDRVRNETLILSDVSDAHLAKEK
IEKIEVNSEVHWEKYTEVGVDGVEIYDITVEGSHNFVANGFIVHTAAVDFEF
TGSWLEAGALVLDGGTALIDELDKMSDRDSVTHEALEQOTISLSKAGITATLNAR
TVVIAAANPKHGFNRKRVSEQIDLPPTLSRFDLIFVLMDEPKVDSEIARHLK
EIKPITAROLEALIRLSEAHARMRLSPIVTRDAREAIKLMETYLKQIAVDEAGOL
DVTILEVGQSARKSLKVERILDIVEKLETSKGAKIDDLLEAAKFGIEKSEAREIL
EKLLKKGQIYMPENGYKTV"

/gene="PH0607"

1630..1992

/gene="PH0607"

1630..1992

/gene="PH0607"

/note="motif=glutamine amidotransferases class-II active
site"

/codon_start=1

/transl_table=11

/product="120aa long hypothetical protein"

/protein_id="BAA29696.1"

/db_xref="GI:3257013"

/translation="MYFLNSCGIFGVTLSPLLTENICIAISLSFSSGSSMRTKI
KSNLLSNVGGRTSCDTLILNLPFGAAITVVLAFVAVMPAFDRILVCCSRAS
WITLLSLILSSSSINA"

3592..3981

/gene="PH0608"

3592..3981

/gene="PH0608"

/codon_start=1

/transl_table=11

/product="129aa long hypothetical protein"

/protein_id="BAA29697.1"

/db_xref="GI:3257014"

/translation="MAETISFLPASSLIFCSSSGDISSSSSLLMGSMVLTSTFLNV
GTFSLFSWTLRPTITLSPGIANVSTISRSIPSNMNLGICPPLKLSGLSLININWF
TNLLFSTTSWTFLLPHCSHFFATKPS"

3693..4040

/gene="PH0609"

3693..4040

/gene="PH0609"

/codon_start=1

/transl_table=11

/product="115aa long hypothetical protein"

/protein_id="BAA29698.1"

/db_xref="GI:3257015"

/translation="MVDLYHLLERWNFLSLFDHSDPHDYPTIWHCCIHVIOEYP
IYEPHPLSSQALWPLILSEHLEVFEFALLHLKLDVFTSALLAFLYCNKALITLRLHN
HMPAILINDEED"

complement(4439..4936)

/gene="PH0610"

gene

CDS

Query Match	11.5%;	Score	227.2;	DB 2:	Length	233000;			
Best Local Similarity	55.4%;	Pred No.	2.6e-59;						
Matches	461;	Conservative	0;	Mismatches	368;	Indels	3;	Gaps	1;
QY	1139	aggaagagaacacagtgcctgctggcgcgaatgacaagttgttcaggagacatacaatatcaa	1198						
DB	228212	ATGGAAAGGAGAAGAGTTACCGACTTAAACAAATGGATAAAGGTGTACACCCTTC	228271						
QY	1199	tacctgaagagattccaatgtctacagaagacggcgctgatggtgaaacgctggcgctgatga	1258						
DB	228272	AACCTGAACACATTTAAGGTTAANGCAAGTGACGGGGTTGAAATAGATGCCTGGGTAATGA	228331						
QY	1259	ggcctgcacaaatggaagtgtagacaacatatccacttattcttaacatacacgccgcgtc	1318						
DB	228332	AACCGGTGAAC TTCAGGAAGAAGAAAGATATCCAGATATCTTAGAGATCCACGGTGGTC	228391						
QY	1319	cgcataatgatgcagacatacatattttcatgatttcaggctgctggcgcgcaaaagat	1378						
DB	228392	CTAAACCGCTTACGGTTACGCTTTTATGACGAGATTCCACGTTTTAACCTCTAAAGGCT	228451						
QY	1379	acgcggctgcttatatacaatccgagagaagaccacgcctacggcaggaatttgtgaatg	1438						
DB	228452	TCTGCTGATPATTTCTCAAATCCTTAGAGGAGCGATGGCTACCGAGAGAGGATTCGCGGAT -	228510						

RESULT	7
AE001879/c	DNA
LOCUS	BCT
DEFINITION	Deinococcus radiodurans R1 section 16 of 229 of the complete chromosome 1.
ACCESSION	AF001879 AF000513
VERSION	AE001879.1 GI:6457832
KEYWORDS	Deinococcus radiodurans.
SOURCE	Deinococcus radiodurans Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus. 1 (bases 1 to 10256)
REFERENCE	White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Mofatt,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen.M., Matathevan,J.J., Lam,P., McDonald,I., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et.al. Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
TITLE	Science 286 (5444), 1571-1577 (1999) 20036896
JOURNAL	2 (bases 1 to 10256)
MEDLINE	White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
REFERENCE	White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
AUTHORS	

TITLE	Genome sequence of the radioresistant bacterium <i>Deinococcus radiodurans</i> R1
JOURNAL	Science 286 (5444), 1571-1577 (1999)
REFERENCE	20036896
AUTHORS	2 (bases 1 to 10256) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Mofat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Nelson, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers 1..10256 "radioresistant-deinococcus radiodurans"

gene
CDS

```
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
/gene="DR0164"
100..1017
/gene="DR0164"
/note="similar to GB:D14846 PID:286165 percent identity:
49.11; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endo alpha-1,4 polygalactosaminidase, putative"
/protein_id="AAF09753.1"
/db_xref="GI:6457833"
/translation="MRGLPTVLLPLSAVLLASCGTSPVAATVFPATPAOPTRTIA
AQATPPLRLPAGLAWDQIAGATEAKVALPAGVSLDLDGFEISAQVADLKAQGV
YTVCLYNGVSEYSPDAQIPDSIKITQDPNPFDESFVDI RDVFRGGVLGILDRR
LALCAAGQDAVEPNLONQONTSGVISRQDQLDFNGMLADRAHAHGLAILOKNGPD
YVLQADQGRMLVDLPDVLNESCORYKEGCLTEYVRRGKLALNVRYRQADLNCATM
NSLGVNALFDLSLVGLKQAYKRVSCGT"
/complement(1079..3046)
/gene="DR0165"
/complement(1079..3046)
/note="similar to GB:A1009126 percent identity: 56.01;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-peptide hydrolase, putative"
/protein_id="AAF09754.1"
/db_xref="GI:6457834"
/translation="MNSETPAPGDSLLALAPPSPQVSPCKQVAFVLAQISREDP
AKPDKFARYRSGLWLSLSEGAARPLTHAETGRGDSAPRPGDGNLAFVBSAGEVK
AALMLPLKGGEARVTHFKNGVSGPQWSPDGRFTAFTTADTEOKDRGERGARVLTR
PYVRAGQDWLPERPAALWLDVDEADKLREWVAPEIGALSNWPSRSLVLLVSEDE
WOASQWQVDYLPPLTADAPAAPOKLLDNNSAAGHLAPHDPGCFALIRPAGKNGT
EHAHLYIENGQHRUDTGHDPVGDAGSDCHVAFPEGRLWDGDTLLFFSTVRGS
VGLFTAHIGGVKAYDHPDQGVLSAFTANEHGVALIRESATREFPELVNGQRTDLHA
RPFPPVREPQRTVFETELGEGWLLPEGEOKFALLNIHGGPHTDYCHGHTFQHL
MAARGVGYCYNRPVSGVQAWDAIYGRWGTVDADLLNFDFCLCAVPRDLDAKT
AVMGSGTGFMTWITGHTTREFQAAITDRCISNLISFGTSDIGLRFWDELGLDFS
RADALKLDLSPLOQVENVKPTPLIVHSVLDHRCVPEAQEQWYALAHKHQVPRVRF
PEENHELRSRGRPRLRLTNEFYAWLERWL"
/3098..3352
/gene="DR0166"
/3098..3352
/gene="DR0166"
/note="similar to SP:P31787 PID:1323021 PID:1620454
PID:170970 percent identity: 66.09; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA-binding protein"
/protein_id="AAF09755.1"
/db_xref="GI:6457835"
/translation="MNFEEAQOQEVNTLSKKPGNDVLLKLYKQSGAGVDGKRPGG
FDFVGAKYDAWAGLKGKTQQAQOQEVVALVEELKAKDQ"
/3663..4649
/gene="DR0167"
/3663..4649
/gene="DR0167"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09762.1"
/db_xref="GI:6457842"
/translation="MPSANVSPCPGVRGGMGPKAKAEASKPHQIPVKLPFVTPAP
DALAAKARMRDLAAYTAAALPGRTHSLMAGVPGVDLKFMPGLWRDGAQDPENHVL
INSAARPERORFTLAHEIGHAILLDGDDLLSDHDAEYEGERLEOVETLCNVAANAAL
MPEVTAELEREFGPTGRALAEALAKRAEYSSASALYALTEQTPVPVIVAYCARGPFR
EQAASDEDGAPSTPEKVLTVRASSTGRVKYTLASGTPVPADHPAALALATAGMEVRES
YVDFRSGRKMAEVDAYPSRGIVAVSFEDPAPRLGRKKSEQADRFQDPAQ"
```

gene
CDS

gene
CDS

```
/gene="DR0169"
5540..5896
/gene="DR0169"
/note="similar to GB:D26185 SP:P28823 GB:M34053 PID:143411
PID:467467 percent identity: 62.18; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydroneopterin aldolase"
/protein_id="AAF09757.1"
/db_xref="GI:6457837"
/translation="MSRVVLEGFHHAHGVFAAEGVLGARFVVDAELHSPFAGIADD
LQAVNYAAAYAAIREEVTEKHQIEVLADRIADRLISDSRLRVRKRVHKKFAPL
PGIFRDVTAETEKRT"
5893..6408
/gene="DR0170"
5893..6408
/note="similar to GB:D26562 SP:P26281 GB:L06495 PID:146013
PID:473801 percent identity: 56.86; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="2-amino-4-hydroxy-6-
hydroxymethylidihydropteridine pyrophosphokinase"
/protein_id="AAF09758.1"
/db_xref="GI:6457838"
/translation="MTRAFIALGANLGDPOATLRRALTELGAEGVRSALYRTAPV
GGPPGQPDYLNVAELETALSAALLAHLAELADAGRTDRVWEARVLDLIVYGE
DVSDDPALTLPHPRWGERGVFLAPLSLAPSLAHPRTGENVAQALRRVLDAGVSESL
SLCWPELGSGE"
/complement(6362..7054)
/gene="DR0171"
/complement(6362..7054)
/gene="DR0171"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09763.1"
/db_xref="GI:6457843"
/translation="MNLEGVHPLRLGLWCMEOTSSSSQOVTDETVASALLDTRSR
IYLSAFLPTPTLSAARDLGRPLNTVKYOMOKLRLGLLKMVCDSPPIYQAGESEFF
IPVELTPAHMPOQMLQLLHTEWNRMDTALLSTADEARRHGVGMQVRKPSNAULTLE
HAASAGWNFTRPQAPVAVLDLGLSELSDFEQAQALQORLLQLLAQYRRKPGSLILR
LTLTASTAKAH"
7034..7849
/gene="DR0172"
7034..7849
/gene="DR0172"
/note="similar to SP:Q50665 PID:1449319 GB:AL123456
percent identity: 49.77; identified by sequence
similarity; putative"
/codon_start=1
```

gene
CDS

```
/gene="DR0172"
7034..7849
/gene="DR0172"
/note="similar to SP:Q50665 PID:1449319 GB:AL123456
percent identity: 49.77; identified by sequence
similarity; putative"
/codon_start=1
```

gene
CDS

```
/gene="DR0172"
7034..7849
/gene="DR0172"
/note="similar to SP:Q50665 PID:1449319 GB:AL123456
percent identity: 49.77; identified by sequence
similarity; putative"
/codon_start=1
```


GRRLRLALRFRAGERWRISMARLVMDLEGVPEGEYFRGDCLCFRGVCIKTYPARG
HLQHTYLLLEPGCLVHLVDITDGFQPMNPEASPELFLDTAEANAMEARGYLSA
KDRVYSRGECVLAARFAGRLVEQAEVYKALEAARGRLPMFWELTGRTIYRYLEGM
EATMSYFEGVMTEKLLGLLGLWGCTRKTPGGYVARDCHILGVKIDRITLSLA"
complement(1196..1684)
/gene="APE1560"
complement(1196..1684)
/gene="APE1560"
/codon_start=1
/transl_table=11
/product="162aa long hypothetical protein"
/protein_id="BA080539.1"
/db_xref="GI:5105246"
/translation="MDKPPGQGVGVPPSGVLNADPPFAEVLPRPELPLRHL
APFQINQSHADPLPPPLAGKPEGKLPKPSGLILRGSHFASPPMPPEPPMWWVEY
SSSHYINIHQPTLPCLTPRTHENLEPRILNQHTVANGKPPPKRKNKSTSPSTAPQ
IT".

```

1L"
2258. .2437
/gene="APES056"
2258. .2437
/gene="APES056"
/note="similar to OWL:SO061050 percent identity:41.667 in
24aa,"
/codon_start=1
/transl_table=11
/product="59aa long hypothetical protein"
/protein_id="BAA80560.1"
/db_xref="GI:5105247"
/translation="MARVIRAVAKKNGPYLVLEVDGKVQVATLCRCGHSNNKPYCDGTHAR
VGQADEKVAFAEF"
2536. .2952

```

```

/gene="APEL561"
2536..2952
/gene="APEL561"
/note="motif=neutral zinc metalloproteinases, zinc-binding
region signature"
/codon_start=1
/trans_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80561.1"
/db_xref="GI:5105248"

```

/translation="MRYERAPDGRCAAEEVLVDALGLDLYTDVSRVYCVRSWGSGRGRAYA
 RITYGTGMAAAALGVNPGCVYEVVSEKFDLPLEEKVKVYIHELLHPITFTSGGLRAH
 GPLVLRGRVAALRYRVRGDARVMGLREALDGGPQE"
 complement(2599, .3063)
 /gene="APE1562"
 complement(2599, .3063)
 /gene="APE1562"
 /codon_start=1
 /transl_table=11
 /product="154aa long hypothetical protein"
 /protein_id="BAA80562.1"
 /db_xref="GI:5105249"
 /translation="MQGSRYVSGFRGAARPGYAGGLGSWFPSPWPSPRCPCWELFLRAS
 QGFADAHNRCITFDPPVQKNCNPVPMQAMGPEAAAGAGYQVEQLYDLDNLLLIQWQ
 RAVELLRLNLINVAGYIAEGRSRSPRPVDPRISSPPPTTPPAHVAHPGVDIV"
 complement(2906, .3376)
 /gene="APE1563"
 complement(2906, .3376)

```

'gene="APE1563"
'codon_start=1
'transl_table=11
'product="156aa long hypothetical protein"
'protein_id="BAA80563.1"
'db_xref="GI:5105250"
'translation="MKKEGVLRLETTTPKGCEDITPDVHAYLGSLGVRETALCYTRV
GRDGALEVLLEGLMTLQAPEPEGEIMAAVTISPGNAILLITQERODGATAASTA
GDAGPEGVVAREGRAGRLRWGLGKLAILAMAGAAALIGLEGLHPRLRGCP"
'complement(3615..4268)
'gene="APE1564"
'complement(3615..4268)
'gene="APE1564"
note="similar to PIR:H70375 percent identity:45.455 in

```

```

ll0aa.motif=C-5 cytosine-specific DNA methylases
signatures"
/codon_start=1
/transl_table=11
/product="217aa long hypothetical protein"
/protein_id="BAA80564.1"
/db_xref="GI:5105251"
/translation="MAGVGGRPPSVGSVVADDTMTDAKLREVLGARSLEAARIM
ARILARASRMLGTVITPSVRAEIERFMGLGNCVHPTTIARLMAWVRKPTTHIELRLP
AAMFRVYTVARLDKGLRAEDYHRALRSKDEGDIVRSLEKRYEYTRKGMILDS
VEVDVDTLLALELGAIAVTSDEGLRRAAEDGLVTLVPVELLEYILALEEVKSVEE"
3888..4304
/gene="APE1565"
3888..4304
/gene="APE1565"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80565.1"
/db_xref="GI:5105252"
/translation="MILSHKPLVPEPGNGLHITPSHGCVWEAOLMRGRLNPHPRHKP
GNCRCVHAHVPQHEPLYLSPDARSYVHKQPIPGSPGQVPSHPYSGLLIQAPGPDLAQLG
VCKHGRCVCHNNAHARRRRPAHASHPLQPPCPSLV"
3955..4299
/gene="APE1566"
3955..4299
/gene="APE1566"
/notes="motif=prokaryotic membrane lipoprotein lipid
attachment site"

```

```

codon_start=1
/transl_table=11
/product="114aa long hypothetical protein"
/protein_id="BAA80556.1"
/db_xref="GI:5105253"
/translation="MAAGRSSCVGGLLTALHIAISAMVVGCTPPPSNNLSITSALTIG
VMYVTSPLERASILAIIILAASSRLLPATSRSLASNVNVYSATTLPTLGGGLPT
PATPSNPVHPFR"
complement(4296) .5405)
/ gene="APF1567"
complement(4296) .5405)
/ gene="APF1567"
note="similar to Swiss_Prot:Q57857 percent
identity:33.430 in 344aa. motif=aminoacyl-transfer
RNA synthetases Class-II motifs"
```

```

'codon_start'=1
'transl_table'=11
'product'='369aa long hypothetical protein'
'protein_id'='BAA80567.1'
'db_xref='GI:510525'
'translation'='MDPQGVSDLEALSTRSVSRSEFDVYVGRQFRCVPGETLVA
RRGEQGLVVMGVPSTIKMLLPVAVPRFPGTFLVLEKMGNGYNVRVTLGGMYAAT
GGLGICPTVTRRLRLYGCAQLILEDAGBSTIAGEVNGLENPVTTRYVSEAPQGF
FTFDLFGKGRQLPPKFTSLAPETGLKTVALLAEIPATASGVELRYIVDELKRGK
GVIMDPGRVEPVPTTSTNIGDILRMGRVYFPEEGSRVLTRELIREFWETGR
VAVRGYDGLRALPALEAVEAVSRGRLVEEPLVANEVEAEVIAFSLGWHLEIA
complement(5525..5854)
name='APEL568'

```

Query Match	8.2%	Score 161.8	DB 2	Length 251700
Best Local Similarity	54.5%	Pred. No. 1.le-38		
Matches 369	Conservative 0	Mismatches 302	Indels 6	Gaps 2
QY 1297	attcttaacatacacggcggtccgcacatatgatgtacgacacatacatattttcatgagttt	1356		
Db 172547	ATTCTCTATATACATGGGGGCCFAAGACGACCTACGGCTTCACACAGGATTC	172606		
QY 1357	caggtgctgcgcgcgaaggatacgcgctgtttatatccagagaggagccacgcgc	1416		
Db 172607	CAAATGCTAGCCTCTCAGGGCTTTGCGGTGATTTACTCGAACCCACAGGGGAGTCACGGT	172666		
QY 1417	tacgcggcaggaatttgaaatgcggtcagagagattatgggggaaggattatgcacgat	1476		

wed Sep 26 11:29:23 2001

Db 208695 TCATGACATAGCATAAGGGGAGGCCAAGACACAGAGTGAAGAGGTACAGCTCTTCAT 208636
QY 1950 ctcagtggttgatcaacatctc 1971
Db 208635 GGAGTCTTCGAGGAGGAGCTC 208614
RESULT 10
AX041921/c 349980 bp DNA PAT 23-NOV-2000
LOCUS Sequence 816 from Patent WO0065062.
DEFINITION AX041921
ACCESSION AX041921
VERSION AX041921.1 GI:11340686
KEYWORDS
SOURCE Pyrococcus abyssi.
ORGANISM Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.
1 (bases 1 to 349980)
Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O.,
Querellou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D.,
Rafin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.
Genome sequence and polypeptides of pyrococcus _i(abissy), fragment
and uses thereof
Patent: WO 0065062-A 816 02-NOV-2000;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER
INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
TITLES Location/Qualifiers
1. 349980
/organism="pyrococcus abyssi"
/db_xref="taxon:29292"
/note="Original length of seq 1: 1765118 (for 812
sequences) New sequences : seq 001, from 0.000.001 to
0.349.980, length: 349.980 seq 813, from 0.300.001 to
0.649.980, length: 349.980 seq 814, from 0.600.001 to
0.949.980, length: 349.980 seq 815, from 0.900.001 to
1.249.980, length: 349.980 seq 816, from 1.200.001 to
1.549.980, length: 349.980 seq 817, from 1.500.001 to
1.765.118, length: 265.118"
BASE COUNT 98084 a 80447 c 77665 g 93784 t
ORIGIN
Query Match 8.0%; Score 158; DB 9; Length 349980;
Best Local Similarity 52.7%; Pred. No. 1.9e-37;
Matches 391; Conservative 0; Mismatches 345; Indels 6; Gaps 2;
QY 1230 cggcgtgatgtgaacggtggtgatgagcctgcaaaatgaagtgaagacaacata 1289
Db 205109 CGACTGGAGATAGACGGCTGGTACATAAAGCCCGAGCTCAAGGAGGCGGAGAGGC--- 205051
QY 1290 tccacttatttcaacataacacggcgtccgcataatgatgtacgacacatacatatttca 1349
Db 205052 TCCGGTAATAGCTCTTCGTCACGGTGGGCCCAAGGAATGTACGGTTACTTCAAGTA 204993
QY 1350 tgaatttcaggtgctggcgaggaagatacggcgtgctttatatcaatccagagaggaag 1409
Db 204992 CGAGATGACAGTTATGGCGAGCAAGGGCTACTACATAGTCTATGTATCAATCCAAAGGGGAG 204933
QY 1410 ccacggctacggcggaggaatttgaaatcggttcagagagattatgggggaaagattaa 1469
Db 204932 CAACGGTTACACGGAGGACTTTGCGACTTACAGTCTTCGAGAGCTGGCTAGAGGACTT 204873
QY 1470 tgaatgtgatgcaggtgctggtgatgaggtcatcaaacagagatccgcattatgatcctaa 1529
Db 204872 CCAGGNATATCTGAATGGAATAGAGAGTCTTCAAGCTTGAACCGGACGACAGGGA 204813
QY 1530 gggcgtcggtgtacagggcggaagctacagaggttttatgaccaactgatgctcgggca 1589
Db 204812 GAGGGTTGGAATAACGGGAATAAGCTACGAGGCTTCATGACGAACACTGGCCCTAACGCA 204753
QY 1590 gacgaaccgctttaaagctgcccgttaccacggcgtcgatatacaaatggatcagctttca 1649

Db 204752 GAGCGACCTATTCAAGCCCGGAATCAGCGAAGCGGAATAAGCTACTGGCTAACGAGTTA 204693
QY 1650 cggcgtcagtgatagctgctatttttttacagagctgcagcttgagcatgacatgtttga 1709
Db 204692 CGCGTTCTCGACATAGGATTTGGTT---CGACAAGAGAGTTATAGGTGAGAACCCGCT 204636
QY 1710 gacacagaaaagctctgggacgggtctcttttaaaataagcagcaaaaacgtggagacacc 1769
Db 204635 AGAAGACAGAACTTCAGGAGCTTAAGTCCGCTCTTCTACGCTAAGAACGCTGAAGGCTCC 204576
QY 1770 gcttttgatactgcatgagcggagcggagtgacccgagcagcagcagcagcagcagcagc 1829
Db 204576 CCTGTTCTAATCCACAGCCTTGAAGATTATCGCTGTCCATTCGATTCAGAGCTTGATGTT 204516
QY 1830 gttatcgctctgaaataatggcaggaacacacagcgttccgctttttccgaatgcac 1889
Db 204515 CTACCACGCTCTGAAAGATTAGGCAAGAAAGTTTACATAGCGATATTCAAGAGGGGAGC 204456
QY 1890 gcaaatattatcagcaccgacacccacacacacacacacacacacacacacacacacacac 1949
Db 204455 TCATGACATAGCATAAGGGGAGGCCAAGACACAGGATGAAGAGGTACAGCTCTTCAT 204396
QY 1950 ctcagtggttgatcaacatctc 1971
Db 204395 GGAGTCTTCGAGGAGGAGCTC 204374
RESULT 11
AP000002 257000 bp DNA BCT 06-APR-2000
LOCUS Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position
DEFINITION (2/7).
ACCESSION AP000002 AB009475 AB009477 AB009478 AB009479 AB009480
AB009481 AB009482 AB009483 AB009484 BA000001
AP000002.1 GI:3236129
VERSION
KEYWORDS Pyrococcus horikoshii (strain:OT3) DNA.
SOURCE Pyrococcus horikoshii
ORGANISM Pyrococcus horikoshii
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.
1 (sites)
Kawarabayasi,Y., Sawada,M., Horikawa,H., Halkawa,Y., Hino,Y.,
Yamamoto,S., Sekine,M., Baba,S., Kosugi,H., Hosoyama,A., Nagai,Y.,
Sakai,M., Ogura,K., Otuka,K., Nakazawa,H., Takamiya,M., Ohfuku,Y.,
Funahashi,T., Tanaka,T., Kudo,H., Yamazaki,J., Kishida,N.,
Oguchi,A., Aoki,K., Nakamura,Y., Robb,T.F., Horikoshi,K.,
Masuchi,Y., Shizuya,H. and Kikuchi,H.
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
DNA Res. 5 (2), 55-76 (1998)
2 (bases 1 to 257000)
Tanaka,T., Kawarabayasi,Y. and Kikuchi,H.
Direct Submission
Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka
Kawarabayasi, National Institute of Technology and Evaluation,
Biotechnology Center; 2chome 49-10 Nishihara, Shibuya-ku, Tokyo
151-0066, Japan (E-mail:genome@3enite.go.jp, Tel:+81-3-3481-8951,
Fax:+81-3-3481-8424)
On Jul 28, 1998 this sequence version replaced gi:3130553
gi:3130577 gi:3130627 gi:3130656 gi:3130685 gi:3130705 gi:3130737
gi:3130785 gi:3130824 gi:3130828.
AB009475-AB009484: submitted (10-DEC-1997)
Kawarabayasi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.
Shizuya, H. is at the California Institute of Technology, Pasadena,

CA, USA.

The other authors are at the National Institute of Technology and Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan. All the sequence with length 100 codons or more between ATG or GTG and stop codon are defined as CDS

Homology analysis is performed by Smith-Waterman algorithm against GenBank and GenPept release 103; EMBL release 52.0; SwissProt release 34.0; PIR-Protein release 54.0; and OWL release 29.5.

E-mail address for comments and questions: genomeOT3@nig.go.jp

Restriction map, ORF organization, sequence alignment and more information are available at W.W.W. site of Biotechnology Center, URL: <http://www.bio.nite.go.jp/>.

FEATURES

Source

1..257000

/organism="Pyrococcus horikoshii"

/strain="On3"

/db_xref="taxon:53953"

/map="287001-544000 nt. position"

/gene="PH0318"

/db_xref="GI:3256710"

/note="similar to PIR:S74842 percent identity: 55.181 in 403aa; PIR:S74842 percent identity: 49.869 in 397aa; Swiss_Prot:P33221 percent identity: 50.773 in 404aa"

/codon_start=1

/transl_table=11

/product="433aa long hypothetical

phosphoribosylglycinamide formyl transferase"

/protein_id="BAA29392.1"

/db_xref="GI:3256709"

/translation="MVVMKLRDELATATDSNAQKILLGSGELGKEIAIEAORLCVGEVADRYANAPAMVAHRSYVGNMDDKFLWSVEREKPDATIIPEIAINLDALEFEFEDGIFVVPNARATWIAHRELRITLTKAKVPTSRMYATVTLDELYEACKEKIGYPCFKAIMSSGKSGSYFYKGPEDIPKAWEEAKTKARGSAEKIIVIEHTDFDVEVTELAVRHFDNGEIVITFPKPGHYQIDGDYHAWOPAEISKAPEVRYIAKRIITDVLUGLGIFGVMEVYKQVWANEVSRPHTDGMVTLASHPGSGSEALHRLVRLGIPGEVWDGYRLPMLIPAAHTVIRAKVSYSPRGLVGLSVNATVRLFGKPEAYVGRRLGLALA WQKDVFAKRAKMAVHAHMLIELTRSDWHDQNYEKRRKHLR"

complement(1645..1974)

/gene="PH0319"

complement(1645..1974)

/gene="PH0319"

/codon_start=1

/transl_table=11

/product="109aa long hypothetical protein"

/protein_id="BAA29393.1"

/db_xref="GI:3256710"

/translation="MFLNYLLCTTSRFLCFPPGFRNVLPGLDVRSLRGAHYSILC VAMITDLFTGLTIEFKGSHVHPGSRNLCFLHKGFPKPLPVHSYPGSSCIWNKYITIL LEFEERI"

complement(2653..3186)

/gene="PH0320"

complement(2653..3186)

/gene="PH0320"

/note="similar to owl:SSOUREK percent identity: 68.000 in 150aa; PIR:S76001 percent identity: 60.694 in 173aa; owl:BAURKEI percent identity: 62.329 in 146aa"

/codon_start=1

/transl_table=11

/product="177aa long hypothetical

phosphoribosylaminimidazole carboxylase catalytic subunit"

/protein_id="BAA29394.1"

/db_xref="GI:3256711"

/translation="MVISMKSEKPLVILMGSDSLPVMKEARILIEERGVPYEITII SAHRTPEAYAKAAEERGIEVIAAGAGAAHLPGIIASLTLPVIGVPIKSKALNG LDSLSTVOMPSPGIPATVAIDNAKNAALLAILGKYPEIKELRLRYMKDKRKEV EKAKRLEMGWEYVLESE"

complement(3275..4036)

/gene="PH0321"

complement(3275..4036)

/gene="PH0321"

/codon_start=1

/transl_table=11

/product="253aa long hypothetical protein"

/protein_id="BAA29395.1"

/db_xref="GI:3256712"

/translation="MISWEIEELDRKIKKSHLILTHEEDASSRCKDILFVILSRK WSLNLMFESISYPLQIIIRILSRFGVDVILKYLENHRILAIVDTFGSHGKATMPGV KYLEGMSSETLPIKYAKAVEDHKKVMDLNLFEGRLYGFAISMSYLVFTPEETL RYLETSAEVRYPAYKKYPRGTNFWLWEGVKDKRLLSVYRADYVVKTRSSLGENG IKRELLVKTTPKPIEELVRPEYEFKGNPKLRVD"

complement(4039..4698)

/gene="PH0322"

complement(4039..4698)

/gene="PH0322"

/codon_start=1

/transl_table=11

/product="219aa long hypothetical protein"

/protein_id="BAA29396.1"

/db_xref="GI:3256713"

/translation="MGIIEELVSSIPGEGILAIIOKELESEGRGLLVKHLLENEN NVFLIYELPRTMTNMENGINVDEYLGENLMIFDFGSMNKIERIKGVYKLSYL DDVYFSKLUKWAITLKTNTIANFWFFTYLSSGICKLFSNPLLYKLIWALREELIK GTRPKTIITYSOLECPILEETVYLASDIIVLETRIINGKVGIIITKGNLIFELFKE V"

4856..6172

/gene="PH0323"

4856..6172

/gene="PH0323"

/note="similar to PIR:A64417 percent identity: 55.478 in 442aa; owl:HUMRPS4 percent identity: 33.738 in 424aa; Swiss_Prot:P12039 percent identity: 33.819 in 352aa; motif="prokaryotic membrane lipoprotein lipid attachment site"

/codon_start=1

/transl_table=11

/product="438aa long hypothetical

phosphoribosylamide--glycine ligase"

/protein_id="BAA29397.1"

/db_xref="GI:3256714"

/translation="MKVLVGGGGRHAIGEALVKGAEIYVVSNNRNPGLMLAKDY GLAKETNVEDVIFKARKWGLTFLAIGPEALFAGIVNALLEEGIPAPPTREARLET NKWAREFMERNILPGKMFRIFFDVOEMRRKWIDEGKVVVVKPLGLTGGKGVVGVY QKDNEEAKYAEHIIIRKDGKVLIEERTDGVETLQVTKVVPIMPLVQDYVPAVE GDVGPITGGMSYSCSNHLLPFTIEGDFERALKTEETIEAMRKEGYPYKGLYGFQM LSGEGPVLIEYNARFGDPEAINVLAVLDDNLLEIAKIGVSGSLRKAFLNKATVVKYI APOQYPODPIKIRIEVDEEIKNEGAKIIVAAVDENITLLGSRALIVGVAUSLEEA ERIAENGVSYVKGPIFYRKDVGTRESVEKRIEMKLGKEFEFNL"

6174..7325

/gene="PH0324"

6174..7325

/gene="PH0324"

/note="similar to PIR:H64316 percent identity: 44.847 in 382aa"

/codon_start=1

/transl_table=11

/product="383aa long hypothetical protein"

/protein_id="BAA29398.1"

/db_xref="GI:3256715"

/translation="MIYMDREELISILEKYDKDKITIGVIGSHSALDADGAKKEGF PVLVAQRHRTYAKYFKRLKDNLTGFIDEVILDKFAQIIEIOELIKRNVI IFNRSFVYTGDKYNEFKVPMFGTSLRTERREKESYVWLLEKAGLPYPEVKP BEIDGVVLMVKLPHAKKRLERGFYTAASYKEFEKSERILBLGVITREDLDKARIE YIGPVNFDFYSPIDEEIILGVDMREFETLDCHVRLPAQQOITLPHQFEPYTV CGHASATRESLEKVFDMAEKYVEATKKYPPGLIGPFTLTQTVVDKOLNFYIFDVAP RTGGTNTHMSMGHPYGNLSRSPMSTGRRTALEIKRGIELDELEKLV"

7392..8486

/gene="PH0325"

7392..8486

/gene="PH0325"

/codon_start=1

/transl_table=11

/product="364aa long hypothetical protein"

/protein_id="BAA29399.1"

RESULT	12					
AR086952						
LOCUS		1869 bp	DNA			PAT
DEFINITION		Sequence 1	from patent	US 5985646.		
ACCESSION		AR086952				

RESULT	13				
AR095647		AR095647	1869 bp	DNA	PAT
LOCUS		Sequence 1	from patent	US 6004796.	
DEFINITION					
					08-SEP-2000

[illegible]

307082 ATCCGAATTTTGGACATGTGCGGTACTAGCGATATAGGCTTTTGGTTCAATGCCATAGA 307141
1689 gctt-----gagcatgacatgtttgaggacacacaaagctctgggacgggtctcc 1739
307142 AAATGCCGCTTAAGGATCCTTGGATATCTGAAATCAAAATTCACATGATGAATGTCACC 307201
1740 tttaaaatacgcagcaaacggtggagacacccgtttttgatactgcatggcgaggggaatga 1799
307202 AATTTTTCACGTTAAGAGAGATTAAACGCCCAATATTGTTTATACACGAGAGAACGATTC 307261
1800 ccgatgccgatcgagcggcgagcagctgtttatcgctctgaaaaaatgggcaagaag 1859
307262 CAGATGTCCTATAGAACAGGCCGCAAAATTTTACACGGCTTTGAAGATAAACGGAGTGA 307321
1860 aaccgaagcttccgttttcgaatgcacgaattatcacgcacgcggacacccaaga 1919
307322 ACAAAACCTTGTCAGAGTATCGGTGATTCACATCAACATGCCGAAAAGGAACCCAGA 307381
1920 acagcggatcaagcgcctgaattatatacagctcagctgattgatacaacal 1968
307382 GAACATGAGAAAGAGGCTAGAACTCAAACTTGAATGGTTCAATCACAT 307430

RESULT 15
AE005114 9825 bp DNA BCT 12-FEB-2001
Halobacterium sp. NRC-1 section 145 of 170 of the complete genome.
ACCESSION AE005114 AE004437
VERSION AE005114.1 GI:10581712
KEYWORDS
SOURCE Halobacterium sp. NRC-1.
ORGANISM Halobacterium sp. NRC-1
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
Halobacterium.

REFERENCE
AUTHORS Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Ballga,N., Thorsson,V., Sbrogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dasarma,S.
From the cover: genome sequence of halobacterium species NRC-1
Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
11016950

2 (bases 1 to 9825)
Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Ballga,N., Thorsson,V., Sbrogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dasarma,S.
Direct Submission
Submitted (14-JUL-2000) Institute for Systems Biology, 4225
Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES
source
1. .9825
/organism="Halobacterium sp. NRC-1"
/strain="NRC-1"
/db_xref="taxon:64091"
262..471
/gene="VNG2299H"
262..471
/gene="VNG2299H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11

gene
CDS

us-09-462-845-1.rge

wed Sep 26 11:29:23 2001

Search completed: September 26, 2001, 08:23:28
Job time: 19340 sec